

الطرق الإحصائية
الأستاذ الدكتور
محمد صبحي أبو صالح
اليازوري
www.yazori.com

الإهـــــداء

إلى أرواح أعزائي الطاهرة، أمي وأبي وزوجتي اعترافاً بفضل والدي بتربيتي وبفضل زوجتي بالوقوف إلى جانبي وتشجيعي .

إلى زوجتي العزيزة لتشجيعها وتضحيتها وصبرها ووقوفها بجانبي .

إلى بناتي وأبنائي الأحباء لصبرهـم وتضحيتهـم .

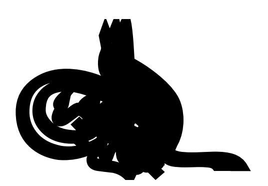
إلى اخوتي وأخواتي وأقاربي لدعمهـم وتشجيعهـم .

إلى أساتذتي وزملائي الأعزاء وإلى طلبتنا الأعزاء الناطقين بلغة الضاد .

إلى كل هؤلاء أهدي جهدي المتواضع داعياً الله العلي القدير أن ينفع أبناء الأمة بما قدمنا .

والله ولي التوفيق

المحتويات



مقدمة

الحمد لله رب العالمين والصلاة والسلام على رسوله الكريم محمد صلى الله عليه وعلى آله وصحبه وسلم الحمد لله الذي هدانا لهذا وما كنا لنهتدي لولا أن هدانا الله. الحمد لله على نعمة الإسلام ونعم من الله لا نحصيها وبعد،

فبين يديك كتاب "الطرق الإحصائية" الذي أردنا من تأليفه رفد المكتبة العربية بأحد الكتب العلمية والذي توخينا في تأليفه الدقة العلمية ووضوح الأسلوب وغايتنا من ذلك كله رضوان الله جل وعلا.

لقد بني هذا الكتاب على منهاج الإحصاء في معظم جامعاتنا الرسمية والأهلية وروعي فيه إمكانية تدريسه في مساقين أو تدريس عدة فصول منه في مساق واحد حسب المنهاج المقرر. هذا بالإضافة إلى كون مادة الكتاب مرجعاً أساسياً للطالب والمعلم والباحث الذي يحتاج إلى استعمال الطرق الإحصائية في التدريس أو تحليل البيانات التي يجمعها عند إجراء البحوث.

إن الغرض من الكتاب تقديم مناقشة وشرح واضح للمفاهيم والطرق الإحصائية والعمليات اللازمة لتطبيقها ، وقد اشتمل على شرح واف للإحصاء الوصفي والإحصاء الاستدلالي وكيفية تطبيقهما.

وقد روعي في تقديم المادة العلمية شرح المفهوم بأسلوب واضح ولغة سليمة ومن ثم إعطاء المعادلات بشكل رياضي دقيق وتبع ذلك أمثلة محلولة لها علاقة وثيقة بالحياة العملية. وفي نهاية كل فصل مجموعة من التمارين التي تكسب الطالب المهارات اللازمة التي تمكنه من تحليل البيانات الإحصائية عل مستوى مادة الكتاب.

لقد احتوى الكتاب على أربعة عشر فصلاً وتألف كل فصل من عدد من البنود ليسهل دراسة المادة العلمية وتدريسها.

تناول الفصل الأول نبذة عن طبيعة الإحصاء والتعريف به، وأنواع البيانات وتدريج القياس والمتغيرات والثوابت وهو يمثل اللبنة الأولى التي تقدم لمادة الكتاب.

أما الفصل الثاني فيعطي طرق عرض البيانات الإحصائية ووصفها كبيانات أولية أو مجمعة ثم عرضها بالطرق البيانية، ويحتوي الفصل الثالث على مقاييس النزعة المركزية والتشتت فيعرفها ويعطي الطرق لحسابها وشرحاً لأهميتها.

ينتقل الفصل الرابع لدراسة الاحتمال وقوانينه. ويشمل الفصل الخامس على المتغيرات العشوائية والتوزيعات الاحتمالية والتوقع الرياضي ودراسة بعض التوزيعات الاحتمالية المنفصلة مثل توزيع ذات الحدين وتوزيع بواسون وغيرها.

أما الفصل السادس فيتعرض لدراسة التوزيعات الاحتمالية المتصلة التي يحتاجها الطالب في الفصول اللاحقة مثل التوزيع الطبيعي وتوزيع t وتوزيع كاي تربيع وتوزيع F. أما العينات وتوزيعات المعاينة فقد بحثت في الفصل السابع بشكل من التفصيل.

اشتمل الفصل الثامن على موضوع التقدير وفترات الثقة مستخدماً موضوعات توزيعات المعاينة.

وتم تقديم اختبار الفرضيات في الفصل التاسع الذي اشتمل على الاختبارات حول الوسط الحسابي والفرق بين وسطين وحول النسبة والفرق بين نسبتين إضافة إلى اختبار الفرضيات حول التباين والنسبة بين تباينين واختبار الاستقلال وحسن المطابقة.

يقدم الفصل العاشر دراسة الارتباط والانحدار بما في ذلك فترات الثقة واختبار الفرضيات حول معامل الارتباط ومعلمات خط الانحدار.

وننتقل في الفصل الحادي عشر لدراسة تحليل التباين بما في ذلك تحليل التباين الأحادي والثنائي والاختبارات البعدية.

يبحث الفصل الثاني عشر يغطي موضوع السلاسل الزمنية ومركبات السلسلة الزمنية ويشمل أمثلة تطبيقية عن بيانات فعلية.

الفصل الثالث عشر موضوع الأرقام القياسية البسيطة والمرجحة واختباراتها.

أما الإحصاءات الحيوية بما في ذلك تعريف بالبيانات الديموغرافية وإحصاءات الوفيات والأمراض وإحصاءات الخصوبة فكانت موضوع الفصل الأخير وهو الفصل الرابع عشر.

ومن هذه المقدمة المختصرة نرى احتواء الكتاب على المواضيع الهامة في الطرق الإحصائية والتي يمكن أن تكون مواضيع مقرر الإحصاء لطلبة كليات العلوم والهندسة والصيدلة والعلوم الطبية، كما أنه يشتمل على مواضيع مقرر الإحصاء لطلبة كليات الاقتصاد والعلوم الإدارية.

إننا نرحب بكل نقد بناء من أساتذتنا الكرام ومن أعزائنا الطلبة ونشكر كل من يرشدنا إلى تحسين مادة الكتاب بتصحيح خطأ أو اقتراح إضافة أو توضيح.

ندعو الله أن يكون عملنا هذا خالصاً لوجهه الكريم وأن نكون قد وفقنا لخدمة أمتنا وأن يكون عملنا هذا من العلم الذي ينتفع به.

المؤلف

محمد صبحي عبد القادر أبو صالح

24 رمضان 1420هـ

الموافق 2000/1/1م

الفصل الأول

طبيعة الإحصاء Nature of Statistics

1:1 مقدمة Introduction

لقد ورد ذكر الإشارة إلى الإحصاء في القرآن الكريم في عدة آيات كريمة، منها: الآية 94 في سورة مريم حيث يقول تعالى "لقد أحصاهم وعدّهم عّدا" صدق الله العظيم. أي علم عددهم وأكّد ذلك تأكيداً.

ومنها الآية 49 في سورة الكهف: يقول تعالى "ووضع الكتابُ فترى المجرمين مشفقين مما فيه ويقولون يا ويلتنا مال هذا الكتاب لا يغادر صغيرة ولا كبيرة إلا أحصاها ووجدوا ما عملوا حاضراً ولا يظلم ربُّك أحداً" صدق الله العظيم. أحصاها هنا: أحاط بها وجمعها وعلمها.

أما الآية 34 من سورة إبراهيم فتقرأ: "وآتكم من كل ما سألتموه وان تعّدوا نعمة الله لا تُحصوها إن الإنسان لظلوم كفار" صدق الله العظيم.

ومعنى لا تحصوها: لا تُطيقوا عدّها ولا تقوموا بحصرها لكثرتها.

ومع أن الآيات الكريمة الآنفة الذكر لم يقصد بها تعليمنا الإحصاء أو معناه أو الفائدة منه ولكننا بكل تأكيد نفهم منها أن الله سبحانه وتعالى قادر على حصر المخلوقات وعدّها والعلم بكل صغيرة وكبيرة متعلقة بها.

أما الآية الأخيرة فتدل على أن الإنسان عاجز عن إجراء المسح الشامل لمجتمع نعم الله ولذلك فلا بدّ له من دراسة عيّنة من هذا المجتمع عندما يريد أن يتفكر في هذه النعم.

ننطلق من هنا لنلاحظ أن للإحصاء علاقةً بدراسة المجتمع ودراسة أجزاء منه ودراسة نشاطاته، وليس المقصود بالمجتمع هنا المجتمع الإنساني بالضرورة ولكن المقصود أي مجموعة من العناصر التي تخضع للدراسة. إذاً فما هو معنى الإحصاء؟

إن الإحصاء، بمعنى الحصر والعد، فكرة قديمة وربما كان قدماء المصريين أول من قام بتطبيقها واستخدامها حيث قام بناة الأهرام بتعداد لسكان مصر وثروتها واستخدموا النتائج في تنظيم مشروع البناء. وفي عصر الدولة الإسلامية استخدم الخليفة المأمون فكرة الحصر والعد لمعرفة عدد السكان ومقدار الزكاة. وكان استخدام الإحصاء في البداية مقصوراً على الأعمال الخاصة بشؤون الدولة كما يدل على ذلك الأصل اللغوي في اسم هذا العلم، وهو بالإنجليزية Statistics وهو مشتق من كلمة State أي الدولة ومعناه "مجموعة الحقائق الخاصة بشؤون الدولة".

يقتصر معنى الإحصاء للشخص العادي على الجداول العددية التي تصف ظاهرة معينة أو على الرسوم البيانية والأشكال التصويرية التي تعرض التغير في ظاهرة خلال فترة زمنية أو في مناطق جغرافية متعددة، وما شابه ذلك.

فكثيراً ما تطالعنا الصحف اليومية بجداول تبين معدل كميات نزول المطر على الأماكن المختلفة في البلد أو معدل كميات نزول المطر على منطقة معينة خلال فترة زمنية محدّدة.

وفي المدارس ترى رسوماً بيانية تظهر أعداد الطلبة في الصفوف المختلفة. أما في مكتب مدير التربية والتعليم لمنطقة معينة فربما ترى جداول تبين أعداد المعلمين وأعداد الطلبة في مدارس تلك المنطقة.

ولو أمعنت النظر لوجدت أن الإحصاء يقوم بوصف طرق متعددة لجمع البيانات والمشاهدات ومن ثمّ يتم تنظيمها وعرضها كما تشاهد في الكراسات الإحصائية التي تصدرها الوزارات والمؤسسات.

أما الإمكانات التي يوفرها الإحصاء للباحثين والطرق العلمية التي يزودهم بها فهي كثيرة ومتعددة وتستعمل هذه الطرق والمفاهيم الإحصائية في كثير من النشاطات، ففي العلوم تمتد التطبيقات الإحصائية من تصميم التجارب إلى تحليل البيانات التي تجمع ومن ثم إلى اختبار الفرضيات أو تقدير بعض المجاهيل المطلوب معرفتها.

أما في الصناعة فهناك التطبيقات القصيرة المدى مثل القرارات العملية اليومية والتحكم، والتطبيقات الطويلة المدى في التخطيط واتخاذ القرارات.

وكثيراً ما تستعمل المؤسسات والشركات الطرق الإحصائية لتحليل نماذج التغير والتنبؤ للنشاطات المستقبلية لتلك المؤسسات أو للاقتصاد بشكل عام. إن كثيراً من هذه التنبؤات يرسي قواعد التخطيط والتحكم.

وبالإضافة للتنبؤ فإن حقولاً مثل التحكم في الإنتاج والتحكم في النوعية تستعمل الطرق الإحصائية كقاعدة أساسية.

وتستعمل المفاهيم الإحصائية في دراسات القوى العاملة واختيار الموظفين وأبحاث التسويق والإعلان والتحكم في المخزون والتجارب الصناعية والتحليل المالي وتدقيق الحسابات وتوظيف رؤوس الأموال والتنمية.

أما في الإدارة العامة والعلوم الاجتماعية فإن الطرق الإحصائية تستعمل بكثرة وخاصة في تحليل القضايا السياسية والاجتماعية.

ويستعمل الإحصاء في إنجاز الدراسات الكمية عن المجتمعات وظاهرة الفقر وحوادث المركبات وفي النماذج الانتخابية والأمور التربوية وقضايا الصحة العامّة كالعلاقة بين التدخين وأنواع متعددة من المرض والعلاقة بين ضغط الدم والعمر والوزن.

إن الباحثين يعتمدون بشكل قوي على دراسة العينات ليحصلوا على البيانات الإحصائية التي تتعلق بنشاطات الإنسان والأحوال المتعلقة به، وأن كثيرا من الطرق الإحصائية قد تمّ استعمالها وتستعمل في الوقت الحاضر لجمع وتحليل وعرض البيانات من مثل هذه الدراسات لغرض التخطيط واتخاذ القرارات.

يلعب التحليل الإحصائي دوراً هاماً في كثير من حقول النشاط الإنساني وهو مفيد جداً في تبادل المعلومات والوصول إلى الاستنتاجات والاستدلالات من البيانات ومن ثم في الإرشاد إلى التخطيط المنطقي واتخاذ القرارات.

1: 2 الإحصاء الوصفي والإحصاء الاستنتاجي (الاستدلالي)

Descriptive and Inferential Statistics.

مما تقدم يتضح جلياً أن هناك نوعين من الإحصاء، الأول هو الإحصاء الوصفي، وهو ذلك الجزء من الإحصاء الذي يعنى بجمع البيانات وتنظيمها وتصنيفها وعرضها عن طريق الجداول أو الرسوم البيانية وغيرها. فعندما يقوم مدير المدرسة بتسجيل عدد الطلبة في كل صف في مدرسته ومن ثم عرض هذه الأعداد على شكل مستطيلات فإنه يستعمل الإحصاء الوصفي في هذه الحالة. وعندما يزور طبيب الأسنان إحدى المدارس ويفحص أسنان كل طالب يضع جدولا يبين فيه عدد الطلبة الذين يعانون من تسوس في أسنانهم ثم يلخص هذا الجدول حسب عدد الأسنان التي أصابها التسوس فيذكر عدد الطلبة الذين يعانون تسوسا في سن واحدة، وعدد الذين يعانون تسوسا في سنين فإنه يستعمل الإحصاء الوصفي في هذه الحالة أيضاً.

أما النوع الثاني من الإحصاء فهو الإحصاء الاستنتاجي، وهو ذلك الجزء من الإحصاء الذي يعنى بتحليل البيانات للتوصل إلى التنبؤ أو الاستقراء واتخاذ القرارات. ويلعب هذا الجزء من الإحصاء دورا هاما في تخطيط التجارب التي تجمع منها البيانات وفي تصميمها.

إذن فالإحصاء يهتم بطرق جمع البيانات وتمثيلها وعرضها (الإحصاء الوصفي) ومن ثم تحليلها وتفسيرها والتوصل إلى الاستنتاجات بناء عليها (الإحصاء الاستنتاجي). والآن ما هو جمع البيانات؟

إنه عملية الحصول على القياسات أو التعدادات أو قيم المشاهدات للتجارب التي يجريها الإحصائي، مستعملا في ذلك مختلف الطرق المتاحة له من إجراء مقابلة شخصية، أو الحديث على الهاتف، أو تعبئة استبانة، أو إجراء القياسات على التجارب الحقلية.

وكلما كان جمع البيانات دقيقا زادت ثقة الدارس في الاعتماد عليها، ولا يكون تحليل البيانات صحيحا ومفيدا إذا كان هناك أخطاء في جمع تلك البيانات.

وبعد أن يتم جمع البيانات يحتاج الباحث إلى "تنظيم وعرض تلك البيانات" أي القيام بعملية وضع البيانات في جداول منسقة وعرضها بطرق مناسبة: أشكال هندسية ورسوم بيانية، وتوزيعات تكرارية.

وبعد ذلك يقوم الباحث بوصف البيانات وذلك بإيجاد قيم عددية لمقاييس مهمة مثل الوسط الحسابي والوسيط والتباين والمدى ونسبة النجاح وغيرها. كل ما سبق يقع ضمن الإحصاء الوصفي، ثم يأتي دور الإحصاء الاستنتاجي الذي يشمل :

أ- "تحليل البيانات" أي إيجاد قيم المقاييس واقترانات معينة تتحدد قيمها من البيانات قيد الدرس وهي البيانات التي حصلنا عليها من العينات فيحسب الباحث الوسط الحسابي للبيانات أو مدى تلك البيانات، أو بعض المقاييس التي تظهر له تباعد البيانات أو تقاربها بعضها من بعض.

ب- "استقراء النتائج واتخاذ القرارات" وهو أهم أهداف علم الإحصاء وأكثرها فائدة حيث يشمل معظم الدراسات الإحصائية والنظريات القائمة عليها والتطبيقات العملية لها. وهو باختصار يتألف من الطرق الإحصائية التي تؤدي إلى الاستنتاجات التي يتوصل إليها الباحث من تحليل البيانات التي حصل عليها من دراسة العينات وحساب بعض المقاييس منها، وغالبا ما تكون على شكل تقديرات أو تنبؤات أو تعميمات أو قرارات رفض أو قبول للفرضيات الإحصائية. ويعتمد الإحصاء على نظرية الاحتمال وتطبيقاتها والنظريات الإحصائية التي بنيت عليها.

وهكذا، تصنف الطرق الإحصائية إلى طرق الإحصاء الوصفي وطرق الإحصاء الاستنتاجي. فالطرق التي تهتم بالبيانات المتوفرة فقط ولا تحاول التعميم من العينة المدروسة إلى مجتمع أكبر هي طرق الإحصاء الوصفي أما المعالجات التي تؤدي إلى تنبؤ أو استنتاج أو تعميم إلى مجموعة كبيرة كان قد تم مشاهدة بعض عناصرها فهي طرق الإحصاء الاستنتاجي.

1 : 3 أنواع البيانات Types of Data

هناك نوعان هامان من البيانات وهما :

1- البيانات النوعية Qualitative or Categorical Data

نحصل على هذا النوع من البيانات عندما تكون السمة (الخاصية) تحت الدراسة هي سمة نوعية والتي يمكن تصنيفها حسب أصناف أو أنواع وليس بقياسات عددية. فمثلا يمكن تصنيف الشعر إلى (أشقر، كستنائي، أحمر، أسود)، وتصنيف لون العيون إلى (أزرق ، عسلي ، أسود ، أخضر)، ونوع الدم إلى (O, A, B, AB).

2- البيانات الكمية أو العددية

Quantitative or Numerical Data

عندما تكون السمة تحت الدراسة قابلة للقياس على مقياس عددي فإن البيانات التي نحصل عليها تتألف من مجموعة من الأعداد وتسمى بيانات كمية أو عددية. والأمثلة على البيانات العددية كثيرة مثل علامات الطلبة في امتحان ما ، عدد أفراد الأسرة، شدة زلزال ، الزمن الذي تنتظره سيارة عند إشارة ضوئية ، درجة حرارة مريض.

1 : 4 تدريج القياس Measurement Scales

أنت في حياتك اليومية، كثيرا ما تقوم بإعطاء قيم عددية للمشاهدات أو الظواهر أو الحالات الاجتماعية، فمثلا أنت تقول: لدى عائلة أحمد خمسة أطفال ذكور وثلاث إناث، أو تقول: كانت علامة سعيد 68 من مائة في أحد الامتحانات، أو كانت درجة حرارة المريض 38.5 درجة مئوية، أو طول قطعة مستقيم 15.2سم. هل هناك أية فروق في المقاييس التي استعملت كل واحد منها في الحالات هذه؟ هل للتدريج على مقياس الحرارة نفس معنى التدريج على المسطرة التي تقيس بها أطوال المستقيمات؟ ما معنى: طول مستقيم صفر سم؟ وما معنى: درجة حرارة قطعة من الجليد صفر درجة مئوية؟ هل تحمل كلمة صفر في الحالتين المعنى ذاته؟

دعنا نحاول الإجابة عن هذه التساؤلات. إن ذلك يتطلب أن نقوم بدراسة تدرج القياس. وسننظر في ذلك إلى أربعة مستويات، هي تلك المستويات التي يختلف فيها تدرج قياس عن تدرج آخر من حيث قابلية المقارنة، وإمكانية المقارنة بين المشاهدات. وهي تتدرج من أدنى مستوى يمكننا مقارنة المشاهدات فيه إلى أعلى مستوى وهذه هي:

(1) التدريج الاسمي Nominal Scale

يستعمل التدريج الاسمي كمقياس لتحديد هوية الأفراد أو العناصر. وبالتالي فهو يصنّف عناصر الظاهرة التي تختلف في النوعية لا في الكميّة. ومن الممكن أن تكون التصنيفات عبارة عن الأنواع المختلفة لظاهرة ما. وكثيرا ما تستعمل الأعداد لتحديد هوية المفردات، وفي هذه الحالة لا يكون للعدد ذلك المدلول الكمي الذي يفهم منه عادة. فمثلا، يمكن استعمال العددين 1, 0 ليدلا على التصنيف حسب

الجنس فيجعل الصفر ليدل على الذكر والعدد 1 ليدل على الأنثى. لاحظ أن العددين 0 ,1 هنا لا يدلان على القيم العددية، ولذلك لا تجري عليها عمليات الجمع أو الطرح أو أخذ المعدل. هذا من جهة، ومن جهة أخرى، فانت تستطيع تعيين أي عددين آخرين ليدلا على الذكر والأنثى.

ومن الممكن أن تكون التصنيفات حسب الفئات أو المسمّيات مثل فئات الوظائف فنقول فئة المعلمين، فئة الأطباء، فئة المهندسين، فئة المقاولين، فئة العمال، أو حسب لون عدسة العين فنقول: زرقاء، سوداء، خضراء، عسلية.

(2) التدريج الترتيبي Ordinal Scale

يقع هذا التدريج في مستوى أعلى من مستوى التدريج الاسمي، فإضافة إلى خواص التدريج الاسمي فإن التدريج الترتيبي يسمح بالمفاضلة، أي ترتيب العناصر حسب سلمّ معين.

وعندما تسجل قياسات العناصر أو المشاهدات حسب التدريج الترتيبي تكون هذه العناصر مرتبة من أعلى إلى أسفل أو بالعكس، مثل: الرتب العسكرية التي تبدأ من رتبة جندي، وتتدرج إلى رتبة لواء أو مشير، والتقدير الذي يحصل عليه خريج الجامعة وهو: مقبول ، جيد ، جيد جداً، وممتاز. وواضح أن التدريج الترتيبي يسمح بالمفاضلة أي بمعرفة أي شخص أو عنصر أطول أو أقصر من الآخر، أكبر أو أصغر، أفضل تقديراً أو أقل تقديراً، وهكذا.

(3) تدريج الفترة Interval Scale

يعطي هذا التدريج معنى للفروق بين المشاهدات، وبذلك يقع في مستوى أعلى من مستوى التدريج الترتيبي. وعليك أن تلاحظ أن وحدات القياس في تدريج الفترة متساوية، وأن الصفر في هذا التدريج رمز اصطلاحي وليس هو الصفر المطلق الذي يعني العدم. ومثال ذلك، التدريج على ميزان الحرارة المئوي، فهو نوع من تدريج الفترة، ووحدة القياس فيه واحدة أي أن المسافة على الميزان بين أي درجتين هي نفسها المسافة بين أي درجتين أخريين. أما الصفر على ميزان الحرارة المئوي فهو اصطلاحي، أي أننا اصطلحنا أن نعين الدرجة صفر مئوية لتدل على درجة تجمد الماء النقي، ولذلك فإن الصفر على ميزان الحرارة المئوي لا يعني العدم وبالتالي فهو ليس الصفر المطلق. ويتضح هذا الأمر عندما نأخذ مثالا آخر على تدريج الفترة، وليكن تدريج ميزان الحرارة الفهرنهايتي. لاحظ أن وحدة القياس في هذا التدريج هي واحدة، أي أنها هي نفسها بين كل درجتين. ولاحظ أيضاً أن الصفر في هذا التدريج صفر اصطلاحي وهو هنـــا

يقابل $(0 - 32) \times \frac{5}{9}$ أي (17.78-) درجة مئوية وذلك من المعادلة $C = F - 32 \times \frac{5}{9}$

بينما يقابل الصفر على ميزان الحرارة المئوي 32 درجة فهرنهايت،

وبالتالي فإن الصفر على كل من هذين التدريجين اصطلاحي. ويظهر تدريج الفترة عدد الوحدات التي يكون فيها شخص أو عنصر أكبر أو اصغر، أكثر أو أقل، من شخص أو عنصر آخر. فإذا كانت درجة حرارة الماء في وعاء هي 80 درجة مئوية، وفي وعاء آخر 40 درجة مئوية فإن الفرق بين درجتي الحرارة هو 40 درجة مئوية، وبالتالي فإنك تستطيع إيجاد الفرق بين القيم على تدريج الفترة، ويكون لهذا الفرق معنى واضح، فنقول في مثالنا السابق: درجة حرارة الماء في الوعاء الأول أعلى من درجة حرارته في الوعاء الثاني بمقدار 40 درجة مئوية. ولكن لا معنى للنسب في تدريج الفترة، ففي مثالنا السابق ليس هناك معنى لقولك: إن سخونة الماء في الوعاء الأول ضعفا (مِثْلا) سخونة الماء في الوعاء الثاني، حيث أنك لا تطيق وضع يدك في الوعاء الأول ولكنك تستطيع ذلك في الوعاء الثاني. إذاً لا معنى هنا

لكلمة ضعفين. والسبب في ذلك أن التدريج على ميزان الحرارة المئوي تدريج فترة، ولا يوجد فيه صفر مطلق ولا يوجد فيه معنى للنسب.

ومن الأمثلة على تدريج الفترة التدريج على مقياس الضغط الجوي (البارومیتر)، وقياس معامل الذكاء، والعلامات التي يحصل عليها الطلبة في الامتحانات.

(4) التدريج النسبي Ratio Scale

يقع هذا التدريج في أعلى مستوى من مستويات التدريج، وأهم فرق بينه وبين تدريج الفترة أن التدريج النسبي يعطي معنى للصفر المطلق، أي أن الصفر على هذا التدريج يعني العدم، وبالتالي فهو ليس صفرا اصطلاحيا بل إن له معنى محدّداً. ومن ثم فإنك تستطيع إعطاء معنى للنسب بين القيم المقاسة بالتدريج النسبي. وأهم الأمثلة على التدريج النسبي هي قياسات الطول والمساحة والحجم والوزن والقوة.

فعلى سبيل المثال: إذا قلنا إن طول النقطة صفر فإن هذا يعني أن طولها معدوم. وبالتالي فهناك معنى للصفر المطلق. وإذا قلنا: طول قطعة مستقيم 30سم وطول قطعة مستقيم أخرى 10سم فإن هناك معنى واضحا لقولنا: طول القطعة الأولى ثلاثة أضعاف طول القطعة الثانية.

1 : 5 المتغيرات والثوابت Variables and Constants

إن من أهداف وغايات البحوث في العلوم الاجتماعية أن نفهم الظواهر الاجتماعية ونفسّرها ونصوغ التنبؤات. ومن أجل ذلك يجب تحول أفكارنا عن هذه الظواهر إلى بيانات فعلية، عن طريق القياس.

وتسمى السمات أو الصفات التي نقيسها على أفراد عيّنة معينة متغيرات. وهكذا فإن المتغير هو تلك السمة، أو الصفة، أو الكمية، التي تتغير قيمتها من عنصر إلى آخر، أو من مشاهدة إلى أخرى. فلو أردت قياس أطوال طلاب أحد الصفوف لحصلت على عدد من القياسات يمثل كل منها طول أحد الطلبة، أي أن الطول متغير. وكمثال آخر، لو سجلت درجات الحرارة في مدينة معّينة، كل يوم، لمدة شهر فإنك ستحصل على عدد من القيم التي تمثل درجات الحرارة، وبالتالي فإن درجة الحرارة في تلك المدينة تعتبر متغَّيرا. إليك أمثلة أخرى على المتغيرات:

1) ضغط الدم لدى الشخص عند قياسه يوميا لمدة شهر.

2) مستوى السكر في الدم لدى الفرد في مجموعة من المرضى.

3) عدد الأطفال لدى كل عائلة في قرية.

والأمثلة على المتغيرات كثيرة جداً، ولا حصر لها، ومنها: كمية نزول المطر اليومي في فصل الشتاء، والصادرات، والواردات في قطر ما على مدى أشهر السنة.

أما الثوابت فهي السّمات والخواص التي لا تتغير وهي تصف ماهية المواد في ظروف معينة مثل: الكثافة النوعية لعنصر ما في ظرف محدد. فمثلاً: الكثافة النوعية للماء النقي في درجة الحرارة العادية هي 1غم لكل سم مكعب. وإليك أمثلة أخرى على الثوابت. فمعامل التمدد لعنصر الحديد النقي ثابت، معامل الاحتكاك بين مادتين محدّدتين ثابت. وهناك ثوابت كثيرة في القوانين الفيزيائية مثل قوة جذب الأرض، ومثل ثابت نفاذية الفراغ للتأثير الكهربائي، العدد الذري لعنصر معين، والعدد الذري للهيدروجين 1، والعدد الذري للذهب 79.

وهناك ثوابت من نوع آخر تقع ضمن اهتمامات الإحصائي أو الباحث في الدراسات الاجتماعية أو الاقتصادية أو الطبية أو الصناعية أو الزراعية وغيرها. وهي تلك الثوابت التي تصف المجتمعات مثل: معدل عمر المصابيح الكهربائية التي ينتجها مصنع معين على مدى عام واحد، أو معدل الدخل السنوي للفرد في قطر معين في سنة معينة. لاحظ هنا أن معدل عمر المصابيح في المصنع المذكور على مدى عام واحد ثابت، لكنه غير معلوم والثوابت التي تصف المجتمع هي مقاييس تتعين قيمتها من المشاهدات على كل أفراد لمجتمع وتسمى "معلمات".

معلمة المجتمع Population Parameter هو الثابت الذي يصف المجتمع وهو عبارة عن مقياس سمة مثل معدل المجتمع.

ولاحظ أيضاً أنك لو أخذت عينّه من مصابيح ذلك المصنع وأردت معرفة عمر الحياة لكل منها لوجدت أن عمر المصباح متغير.

فأنت ستحصل على عمر معين لكل مصباح، ويمكن أن يتغير هذا العمر من مصباح لآخر، وكذلك معدل أعمار المصابيح التي أخذتها في العينة، فهو متغير أيضاً وسيأخذ قيما متعددة تبعا للعينات المختلفة التي تأخذها من المصابيح لقياس أعمارها. إن مثل هذه المتغيرات تسمى إحصاءات.

الإحصاء Statistic هو متغير تتحدد قيمته من العينة، أي أنه مقياس سمة نجد قيمته لأفراد العينة فقط. فمثلا معدل أعمار (طول حياة) عينة من المصابيح الكهربائية هو إحصاء، حيث السمة التي نقيسها هي عمر المصباح ونقيس هذه السمة لأفراد عينة فقط ثم نجد معدل هذه الأعمار.

1 : 6 المتغيرات المنفصلة والمتصلة Discrete and Continuous Variables

عرّفنا المتغير في القسم السابق وقلنا إنه يأخذ قيما مختلفة تتغير من عنصر إلى آخر وتسمى مجموعة القيم التي يمكن أن يأخذها متغير ما "مجال المتغير". فمثلا: يمكن أن يأخذ معدل الطالب الناجح في امتحان شهادة الدراسة الثانوية أي قيمة مقرّبة إلى خانة عشرية واحدة وواقعة ما بين 50.0 إلى 100.0، وبالتالي فإن مجال هذا المتغير يكون المجموعة

{50.0, 50.1, 50.2, ... , 99.8, 99.9, 100.0}

ويمكن تصنيف المتغيرات تبعا لمجالها، أي مجموعة القيم التي يمكن أن تأخذها. وتقسم المتغيرات حسب هذا التصنيف إلى نوعين هما المتغير المنفصل والمتغير المتصل.

(1) المتغير المنفصل Discrete Variable

المتغير المنفصل هو ذلك المتغير الذي يأخذ قيما قابلة للعدّ أي أنها تكون محدودة أو لا نهائية معدودة وفي كثير من الأحيان تكون القيم التي يمكن أن يأخذها المتغير أعداداً صحيحة، وتمثل عدد الوحدات لكل عنصر أو مشاهدة.

مثال (1)

عدد الأطفال لدى العائلة متغير منفصل لأن القيم التي يمكن أن يأخذها هذا المتغير هي 0, 1, 2, 3,..... وهكذا.

فإذا لم يكن هناك أي عائلة لديها أكثر من 10 أطفال، فإن القيم التي يمكن أن يأخذها المتغير المذكور هي المجموعة {10 , ... ,0, 1, 2}وهذه عناصر يمكن عدّها. وبالتالي فالمتغير منفصل.

مثال (2)

المعدل المئوي للطالب الناجح في شهادة الدراسة الثانوية مقرّبا إلى خانة عشرية واحدة، متغير منفصل، ومجموعة القيم التي يمكن أن يأخذها هي

{50.0, 50.1, 50.2, ... , 99.7, 99.8, 99.9, 100.0}

لاحظ أن بإمكانك عدّ عناصر مجال المتغير المنفصل.

وهناك خاصية هامة للمتغير المنفصل، هي وجود قفزات بين القيم التي يمكن أن يأخذها، ففي مثال (1) كانت القفزات أعداداً صحيحة موجبة وفي المثال (2) كانت القفزات بمقدار 0.1 فإذا كان معدل أحد الطلبة 80.1 ومعدل طالب آخر 80.2 فمعنى هذا أنه لا يوجد أي معدّل بينهما، أي أنه ليس هناك معدل قيمته 80.17 مثلا.

مثال (3)

عدد المرضى الذين يتم إدخالهم إلى المستشفى في اليوم متغير منفصل، لأن عددهم أما أن يكون 0 أو 1 أو 2 أو 3، وهكذا. وبالتالي فإنه يمكن عدّ عناصر المجال، كما توجد قفزات بين عناصره.

(2) المتغير المتصل Continuous Variable

إذا كان مجال المتغير فترة أو عدة فترات سمي المتغير متغيرا متصلا، وليس هناك قفزات بين قيم المتغير المتصل. من الأمثلة على المتغير المتصل درجة الحرارة، والوزن، والطول، والزمن، والعمر وغيرها.

لاحظ أنه عندما نقول: درجة الحرارة متغير متصل، فنحن نعني أنه لو كانت درجة الحرارة 20 درجة مئوية ثم ارتفعت إلى 24 درجة مئوية فإن ارتفاعها يتم بشكل متصل، وبدون قفزات من 20 إلى 24، وبالتالي فإن مجال هذا المتغير (درجة الحرارة) هو الفترة [24 ، 20].

تلاحظ أيضاً أننا عندما نقيس درجة الحرارة نقيسها بشكل تقريبي، وعلى درجة معينة من الدقة. فإذا كان التقريب لخانة عشرية واحدة وقسنا درجة الحرارة في لحظة معنية فإننا ربما نجدها 20.3 أو 21.2 أو 20.7 وهكذا.

كذلك تلاحظ أنك تحصل على قيمة المتغير المتصل بالقياس لا بالعدّ، فمثلا: الزمن متغير متصل، ونحن نقيسه بواسطة الساعة، وبشكل تقريبي، وعلى درجة معينة من الدقة قد تصل إلى واحد بالمائة من الثانية أو أكثر من ذلك.

مثال (4)

إذا كانت سعة خزان ماء 1000 لتراً وبدأنا نصب ماء في هذا الخزان، فما هو المتغير؟ وما هو مجاله ، وما نوعه ؟

الحل:

المتغير هو حجم الماء الموجود في الخزان في أي لحظة. من الواضح أن هذا المتغير متصل، لأنه يقيس الحجم، وهو يأخذ أي قيمة من 0 لترا إلى 1000 لترا. وعند قياس قيمة هذا المتغير في لحظة معينة نحتاج إلى قياس حجم الماء الموجود في الخزان ونسجل القيمة مقربة إلى خانة عشرية واحدة أو خانتين عشريتين حسب دقة الجهاز الذي نقيس به وتكون القيمة واقعة في الفترة من 0 إلى 1000.

مثال (5)

طول الطالب في فصل معين متغير متصل. من الواضح أن طول الطالب متغير متصل، لأنه ليس هناك قفزات عند قياس الطول. فمهما كان طول أحد الطلبة قريبا من طول طالب آخر فإن هناك، من وجهة نظرية، طالبا يقع طوله ما بين طولي الطالبين الأولين، وبالتالي فإن الطول متغير متصل.

ولا بدّ من ملاحظة أن قياس الطول عمليا يكون بشكل تقريبي ويعتمد على الدقة التي يمكن الحصول عليها باستعمال أداة القياس المتاحة فإذا كانت دقة أداة القياس لأقرب مليمتر فإن القيم التي يمكن أن نحصل عليها عند قياس أطوال الطلبة تكون مثلا 159.7سم أو 172.3سم وهكذا.

1 : 7 المجتمع والعيّنة The Population and the Sample

إن من مراحل البحث العلمي تحديد مشكلة البحث وأبعاده وأهدافه وتصميم منهج البحث. وتتطلب هذه المرحلة من الباحث تعريف المجتمع الإحصائي تعريفا دقيقا.

مم يتكون المجتمع الإحصائي؟ إنه يتكون من مجموعة العناصر أو جميع الأفراد الذين ينطوون تحت لواء الدراسة الإحصائية ويهدف تعريف المجتمع الإحصائي إلى تعيين الحدود الصريحة لعملية جمع البيانات ولعملية الاستقراء والاستنتاجات التي يمكن الحصول عليها من الدراسة.

ويمكن أن تكون عناصر المجتمع أفرادا أو عائلات أو موظفين أو مدارس أو مؤسسات أو غير ذلك. لكنه يجب أن يكون ذلك المجتمع، معرّفا تماما ومحددا بحدود الزمن والفراغ (الطبيعي والجغرافي) بحيث يستطيع الباحث معرفة انتماء أي عنصر لذلك المجتمع أو عدم انتمائه. وهذا ما يتم العمل عليه في أعمال المسح، وتسمى كشوف أعمال المسح التي تسجل فيها جميع عناصر المجتمع "الإطار" أو "إطار المعاينة". أما "التعداد" فهو المسح الذي يحاول أن يضم كل عنصر في المجتمع. ونحن نقول "يحاول" لأن الباحث في المجتمعات الكبيرة جدا لن يستطيع ضم جميع الأفراد في المجتمع تحت الدراسة، أما العينة فهي مجموعة جزئية من المجتمع، وسنأتي لدراسة العينات في الفصل السابع.

التمارين

1-1 : صنف العبارات التالية بأنها تنتمي للإحصاء الوصفي أو الإحصاء الاستقرائي.

(أ) من بين 100 شخص أخذوا المطعوم ضد السحايا في مركز صحي، فإن 17 شخصاً حدث معهم مضاعفات.

(ب) لدى زيارة طبيب الأسنان لإحدى المدارس وجد أن 14% من طلبة تلك المدرسة لديهم تسوس في الأسنان.

(جـ) بلغت خسارة مكتبة احدى الجامعات من جراء إتلاف الكتب أو فقدانها خلال العام الجامعي الحالي 3500 دينار أردني فوضعت المكتبة في ميزانيتها للعام الجامعي القادم ما بين 3200 ، 3800 دينار أردني لمعالجة الخسارة من إتلاف الكتب وفقدانها.

(د) نتيجة لاستطلاع حول أزمة المياه قامت به إحدى الصحف اليومية والذي شمل 1000 شخص كتبت الصحيفة أن معظم الأردنيين يخشون من حدوث أزمة ماء هذا الصيف.

(هـ) كتبت إحدى الصحف أن منتخبنا لكرة القدم لعب 8 مباريات دولية، ربح في اثنتين وتعادل في اثنتين وخسر في أربعة.

(و) بلغت مخالفات تجاوز السرعة المسموح بها على خط سريع 42 في يوم ما، فكتب الشرطي المسؤول في تقريره أن معدل عدد مخالفات تجاوز السرعة هو 42 في اليوم وبالتالي فهو 1260 في الشهر.

2-1 : بين نوع البيانات التالية :

(أ) لون عيون أفراد عائلة عددهم 7.

(ب) محافظات إحدى الدول.

(جـ) رواتب 20 معلما في مدرسة ثانوية.

(د) نوع الدم عند الإنسان.

(هـ) عدد الساعات التي يقضيها طالب لمدة 30 يوميا للتحضير إلى امتحان عام.

(و) المصروف الشهري لطالب جامعي.

3-1 : صنف المتغيرات حسب نوعها : منفصل أو متصل :

(أ) عدد السيارات المباعة لدى احدى وكالات السيارات في الأسبوع.

(ب) وزن الطفل عند الولادة.

(جـ) مدة مكالمة هاتفية.

(د) حجم الدم الذي يخسره مريض عند إجراء عملية جراحية.

(هـ) عدد الودائع الجديدة لدى أحد البنوك في الأسبوع.

(و) عدد المساقات التي يسجلها طالب جامعي في فصل معين.

(ز) ضغط الدم لدى مريض.

4-1 : أذكر التدريج المستعمل في كل مما يأتي :

(أ) الترمومیتر المئوي.

(ب) المسطرة حيث التدريج عليها بالسم والانش.

(جـ) البارومتر لقياس الضغط الجوي.

(د) الرتب الأكاديمية لأعضاء هيئة التدريس في الجامعات.

(هـ) تسجيل ألوان الطيف الشمسي.

(و) تسجيل ألوان العيون.

الفصل الثاني
عرض البيانات الإحصائية ووصفها.

Presentation and Description of Data

1:2 طرق عرض البيانات Methods of Presenting Data

تواجهنا في الحياة العملية كميات كبيرة من البيانات، منها ما هو خاص بالوزارات والمؤسسات ومنها ما يتعلق بنتائج التجارب في العلوم المسلكية والعلوم الطبيعية والزراعة وغيرها. فإذا ما عرضنا هذه البيانات بطريقة المقال ضمن التقارير أو الصحف اليومية فإنها بلا شك تكون مملة ويصعب استيعابها والمقارنة بين مفرداتها ولذا كان من الضروري عرض هذه البيانات بطرق شيّقة سهلة، ومن هذه الطرق:

(1) طريقة الجداول Tables

وهي عبارة عن وضع البيانات في جداول وكثيراً ما تستعمل في عرض تغير ظاهرة مع الزمن أو مع مسمّيات كالبلدان والمدارس وغيرها أو مع الزمن والمسمّيات معاً. وعند استعمال هذه الطريقة يجب مراعاة ذكر ما يأتي:

أ- عنوان الجدول.

ب- الوحدات المستعملة.

جـ- مذكرات المصادر التي أخذت منها البيانات.

د- مذكرات تفسيرية تفسر سبب شذوذ بعض البيانات إن وجدت.

مثال (1):

كان عدد الطلبة في إحدى المدارس الأساسية في سنة 1999 كما في الجدول (1)

الجدول (1)

الصف	عدد الطلبة
الأول	45
الثاني	40
الثالث	40
الرابع	36
الخامس	32
السادس	30
السابع	30
الثامن	25
التاسع	25
العاشر	25

لاحظ أن هذا الجدول من نوع عرض تغير ظاهرة مع مسميات، حيث الظاهرة هي عدد الطلبة والمسمّيات هي:

الصفوف العشرة (من الأول حتى العاشر) التي تتألف منها المدرسة.

(2) طريقة المستطيلات أو الأعمدة Bar Graph

تتلخص هذه الطريقة بوضع المسمّيات على محور أفقي ورسم مستطيل على كل مسمى يكون طول ارتفاعه ممثلاً للقيمة المقابلة لذلك المسمى وذلك باستعمال مقياس رسم مناسب.

ونستعمل هذه الطريقة لعرض تغير ظاهرة مع الزمن أو مع مسمّيات أو كليهما معاً، حيث يمكن استعمالها للمقارنة بين قيم الظواهر حسب المسمّيات على مدى عدة سنوات، كأن تقارن بين أعداد الطلبة حسب تخصصاتهم في الجامعة على مدى ست سنوات، أو نعرض أعداد الحاصلين

على شهادة الدراسة الثانوية في سنة معينة حسب المحافظات.

مثال (2): يمثل الجدول (2) أعداد الطلبة في إحدى الكليات في جامعة خاصة خلال السنوات 98/1999 - 94/1995

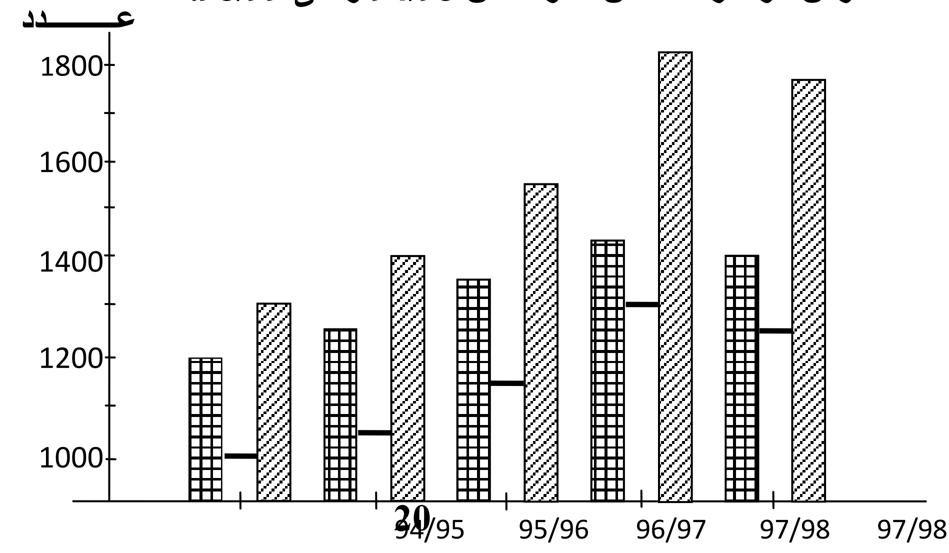
الجدول (2)

المجموع	الإناث	الذكور	السنة
800	200	600	94 / 95
1000	300	700	95 / 96
1300	450	850	96 / 97
1850	800	1050	97 / 98
1750	750	1000	98 / 99

أعرض هذه البيانات بطريقة المستطيلات.

الحل: نعرض هذه البيانات بطريقة المستطيلات كما يظهر في الشكل (1) حيث نمثل السنوات على الخط الأفقي ونرسم مقابل كل سنة ثلاثة مستطيلات، الأول مخطط بخطوط متقاطعة ويمثل الذكور والثاني أبيض ويمثل الإناث والثالث مخطط بخطوط مائلة ويمثل المجموع، على أن يكون طول ارتفاع كل مستطيل متناسباً مع عدد الطلاب الذي يمثله حسب مقياس رسم مناسب.

لاحظ أن هذا المثال يمثل عرض عدة ظواهر مع الزمن حيث هناك ثلاث ظواهر هي عدد الطلاب الذكور، وعدد الطالبات والمجموع، بينما الزمن هو فترة الخمس سنوات من 94/95 وحتى 98/99.



الشكل (1)

: يمثل الطلاب الذكور. : يمثل الطالبات. : يمثل المجموع.

(3) طريقة الخط المنكسر Broken Line Graph

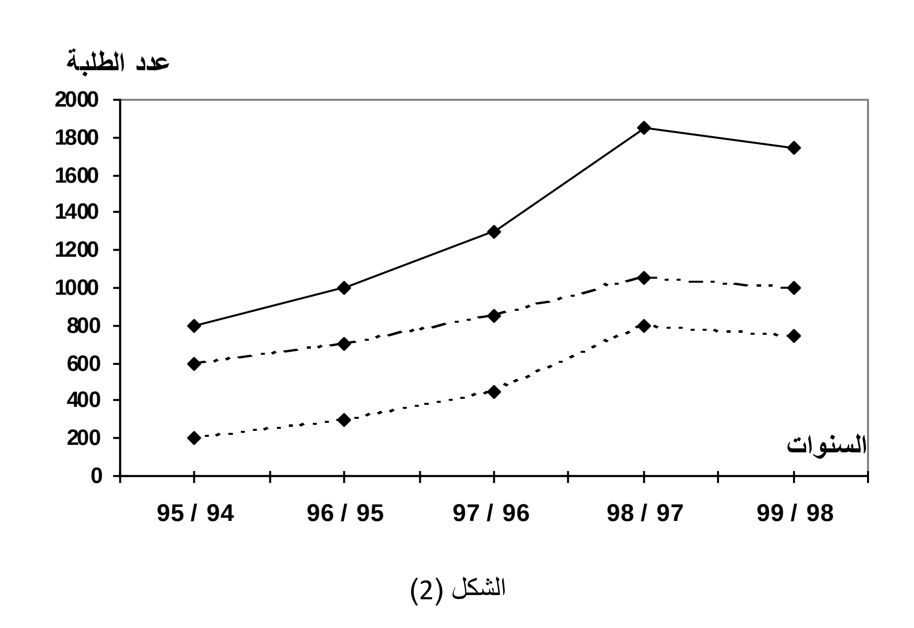
تستعمل هذه الطريقة لعرض البيانات الناتجة من تغير ظاهرة أو عدة ظواهر مع مسمّيات أو مع الزمن أو كليهما مثل تغيّر درجة حرارة مريض مع الزمن بالساعات، أو تغير أعداد الطلاب في جامعة مع السنوات، أو تغير أعداد الطلاب حسب الكليات على مدى فترة زمنية محدّدة.

مثال (3)

إعرض البيانات في المثال (2) بطريقة الخط المنكسر.

الحل: لعرض هذه البيانات بطريقة الخط المنكسر، نرسم محورين متعامدين يمثل المحور الأفقي السنوات والمحور العمودي يمثل أعداد الطلاب بمقياس رسم مناسب.

نعرض أولاً عدد الذكور مقابل السنوات ويمثله الخط المنكسر (   ) ثم عدد الإناث مقابل السنوات ويمثله الخط المنكسر (   ) ثم المجموع مقابل السنوات ويمثله الخط المنكسر (-) ويرسم كل خط برصد النقاط التي احداثياتها الأولى السنوات واحداثياتها الثانية هي أعداد الطلبة، كما في الشكل (2).



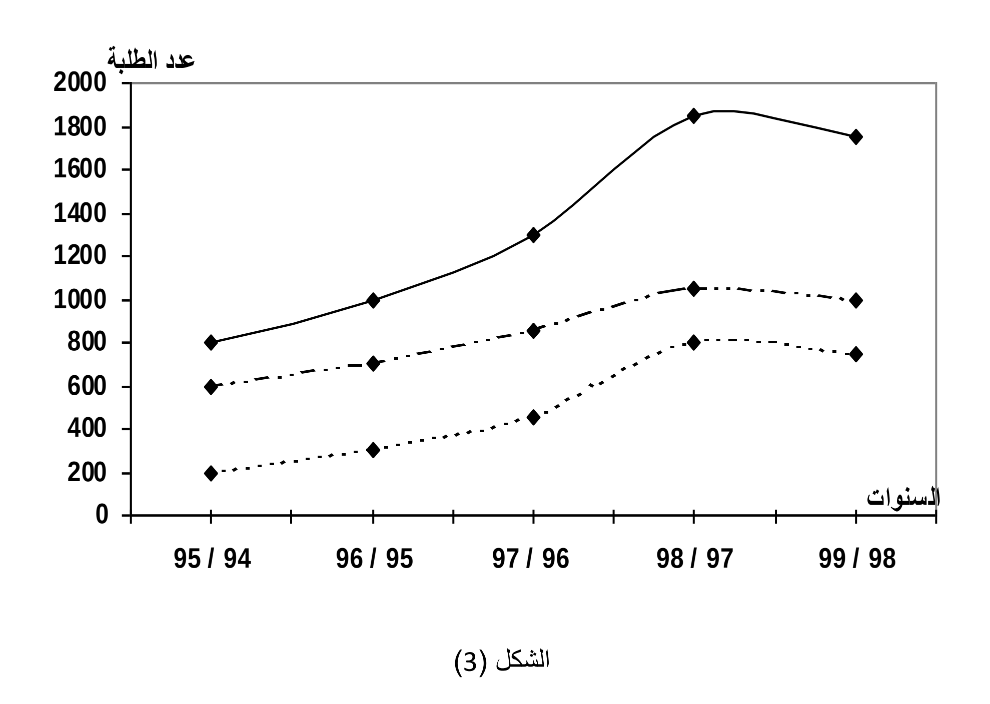
الشكل (2)

(4) طريقة الخط المنحني Curve

وهذه الطريقة تماثل طريقة الخط المنكسر ونحصل عليها بتمهيد الخط المنكسر ليصبح على شكل منحنى بدون زوايا وتستعمل هذه الطريقة عندما تتغير الظاهرة على فترات زمنية قصيرة وكثيرة.

مثال (4) :أعرض البيانات في المثال (2) بطريقة المنحنى :

الحل : نمهد الخط المنكسر الممثل لعرض أعداد الطلاب الذكور في شكل (2) ليصبح منحنى. وكذلك نمهد الخط المنكسر الممثل لعرض أعداد الطالبات وذلك الممثل لعرض المجموع كما يظهر في الشكل (3).



الشكل (3)

1-5 طريقة الدائرة Pie Chart

وأهم استعمالات هذه الطريقة يكون بتقسيم الكل إلى أجزائه، فيمثل المجموع الكلي بدائرة كاملة ويمثل كل جزء بقطاع دائرة يكون قياس زاويته مساوياً 360^0 مضروباً في نسبة الجزء للمجموع الكلي.

مثال (5) : يمثل الجدول (3) عدد أعضاء هيئة التدريس في إحدى الجامعات خلال السنوات 98/1999 - 95/1996

الجدول (3)

العام الجامعي	عدد أعضاء هيئة التدريس
95/1996	90
96/1997	105
97/1998	120

135	98/1999

اعرض هذه البيانات بطريقة الدائرة.

الحل:المجموع الكلي لعدد أعضاء هيئة التدريس =

$$90 + 105 + 120 + 135 = 450$$

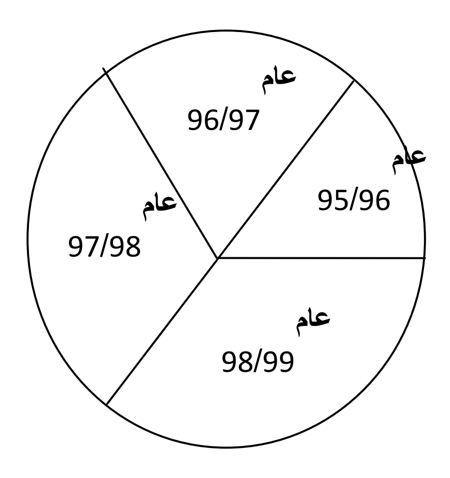
قياس زاوية قطاع 95/1996 هو $\frac{90}{450} \times 360^0 = 72^0$

قياس زاوية قطاع 96/1997 هو $\frac{105}{450} \times 360^0 = 84^0$

قياس زاوية قطاع 97/1998 هو $\frac{120}{450} \times 360^0 = 96^0$

قياس زاوية قطاع 98/1999 هو $\frac{135}{450} \times 360^0 = 108^0$

نرسم دائرة ونرسم القطاعات الأربعة التي تمثل السنوات حسب قياس زاوية كل قطاع كما يظهر في الشكل (4).



شكل (4)

عدد أعضاء هيئـة التـدريس

في إحدى الجامعات للأعوام 95/1996 - 98/1999

(6) طريقة الصور

وتستعمل هذه الطريقة لعرض البيانات بصورة مبسطة مشوّقة كما هو الحال في التقارير الحكومية وكتب علم النفس وفي كتب الأطفال والدعاية. فإذا أردنا عرض البيانات المتعلقة بقيمة الودائع السنوية في عدد من المصارف (البنوك) فإننا نرسم صورة كيس نقود واحد ليمثل كل 20 مليون دينار أردني فإذا بلغت الودائع في البنك أ قيمة 100 مليون دينار فإننا نرسم خمسة أكياس لتمثل هذا المبلغ، وإذا كانت الودائع في البنك ب ما قيمته 60 مليون دينار نرسم صورة ثلاثة أكياس مقابل هذا البنك وإذا بلغت الودائع في البنك جـ 70 مليون دينار فإننا نرسم ثلاثة أكياس ونصف الكيس مقابل ذلك البنك. وكما يلاحظ فإن هذه الطريقة ليست دقيقة جداً.

2 : 2 التوزيع التكراري Frequency Distribution

لدى عرض كثير من الظواهر بالطرق السابقة، لا بدّ وأنك تلاحظ أن هذه الظواهر ربما تحتوي على أعداد كبيرة من البيانات ولذلك فعند عرضها بطريقة الجداول أو المستطيلات مثلاً فإنك لا تستطيع المقارنة بين مفردات هذه البيانات ويصعب عليك فهمها، ولذلك كان لا بد من تلخيصها وعرضها بطريقة مبسّطة تسهل عليك فهمها ومن هذه الطرق التوزيعات التكرارية.

إن التوزيعات التكرارية هي إحدى هذه الطرق التي نتمكن بواسطتها من تنظيم البيانات الكثيرة بحيث لا تخسر هذه البيانات من أهميتها إلا الشيء اليسير أو ربما لا تخسر شيئاً.

أما الطريقة الأساسية لبناء التوزيع التكراري فهي عبارة عن تقسيم مدى قيم البيانات إلى فئات وحصر عدد البيانات الواقعة ضمن كل فئة.

مثال (6):تبين البيانات التالية علامات 25 طالباً في امتحان قصير.

5	7	8	5	6
8	6	8	9	8
7	9	7	10	5
10	8	9	3	7
6	5	7	7	3

اعرض هذه البيانات في توزيع تكراري.

الحل: نلاحظ أن مدى هذه البيانات (أعلى علامة ناقصاً أدنى علامة) صغير وبالتالي فيمكننا أن نضع هذه البيانات في التوزيع التكراري التالي:

الجدول (4)

التوزيع التكراري لعلامات الطلبة.

التكرار	العلامة
2	3
0	4
4	5
3	6
6	7

8	5
9	3
10	2
المجموع	25

ونلاحظ في بناء هذا التوزيع أننا بدأنا من أدنى قيمة وهي 3 ثم رتّبنا القيم ترتيباً تصاعدياً إلى أن وصلنا أعلى قيمة 10 كما يظهر في العمود الأول.

أما عناصر العمود الثاني فتمثل عدد المرات التي تكررت فيها كل قيمة، أما القيمة التي لم تظهر في البيانات فتكرارها صفر.

في هذا المثال، يتبين لدينا أن التوزيع التكراري هو عبارة عن جدول يتألف من:

1- فئات قيم المشاهدات أو القياسات.

2- التكرارات المقابلة لكل فئة أو قياس.

وعند بناء التوزيع التكراري يجب مراعاة النقاط التالية:

1- يجب أن تكون الفئات منفصلة عن بعضها البعض (غير متداخلة فيما بينها).

2- أن تكون الفئات متساوية في الطول ما أمكن.

3- يجب أن تكون الفئات كافية لاحتواء جميع البيانات.

وهذا يعني أنه إذا اخترنا أي قيمة في البيانات فإننا سنتمكن من وضعها في فئة واحدة فقط.

وبهذا نتمكن من إفراغ جميع البيانات في فئات التوزيع التكراري وبدون التباس وسيكون مجموع التكرارات مساوياً لعدد البيانات.

2 : 3 بناء التوزيع التكراري

إذا كان مدى البيانات صغيراً أمكن بناء التوزيع التكراري مباشرة كما في المثال (6). أما إذا كان المدى كبيراً أو كان عدد البيانات كبيراً فإنه يجدر في هذه الحالة أن تقسم قيم البيانات إلى فئات يتراوح عددها ما بين 5، 15 فئة حسب كون عدد البيانات صغيراً أو كبيراً. وبعد تقسيم قيم البيانات إلى فئات تفرغ البيانات على الفئات وتجمع التكرارات المقابلة لكل فئة. نشرح الخطوات المتبعة لبناء التوزيع التكراري بدراسة المثال.

مثال (7):

تمثل البيانات التالية علامات 60 طالباً في الامتحان النـهائي لأحد المساقات.

25	48	32	27	48	44
23	46	33	29	42	24
27	49	23	23	46	25
28	36	41	48	37	25
31	41	43	47	39	36
35	43	48	33	41	43
28	38	46	36	36	48
27	32	34	22	28	47
41	24	24	39	33	38
44	45	46	44	23	43

ولعرض هذه البيانات في توزيع تكراري ذي فئات متساوية نتبع الخطوات التالية:

1- نأخذ فئات متساوية ونفرض أن عددها 6

2- المدى للبيانات = أكبر قيمة - أصغر قيمة

من البيانات التي لدينا نجد:

المدى = 27 = 22 - 49

3- نقرر طول الفئة C وذلك بقسمة المدى على عدد الفئات ثم نقرب الجواب دائماً إلى أعلى بحيث تكون دقة طول الفئة تساوي أو تقل عن دقة الأرقام المستعملة في البيانات.

$$\text{ففي مثالنا يكون: } \frac{\text{المدى}}{\text{عدد الفئات}} = \frac{27}{6} = 4.5$$

وبما أن البيانات معطاة لأقرب عدد صحيح (لا يوجد كسور عشرية) نقرب العدد 4.5 إلى أعلى فنأخذ طول الفئة C = 5

يقصد بدرجة الدقة التي أعطيت بها البيانات، أنه إذا كانت البيانات أعداداً صحيحة فإن ذلك يعني أن طول الفئة يجب أن يكون عدداً صحيحاً، وإذا احتوت البيانات على خانة عشرية واحدة فيجب أن يكون طول الفئة عدداً صحيحاً أو ذا خانة عشرية واحدة، أي أن طول الفئة لا تزيد دقته عن الدقة المستعملة في البيانات.

4- نعين الحد الأدنى لأول فئة ويجب أن يكون هذا الحد مساوياً لأصغر قيمة في البيانات أو أقل منها بقليل، ويجب وأن تكون درجة دقته نفس درجة دقة البيانات المستعملة، وإذن فالحد الأدنى للفئة الأولى هو 22 وبعد تعيين الحد الأدنى للفئة الأولى نعين الحد الأدنى الفعلي لتلك الفئة وهو عبارة عن الحد الأدنى ناقصاً نصف وحدة دقة أي نصف وحدة من الوحدات التي قربت إليها الأعداد في البيانات. وإذن فالحد الأدنى الفعلي للفئة الأولى هو: 21.5 = 0.5 - 22 لأن التقريب في البيانات المستعملة كان لأقرب ساعة (لأقرب عدد صحيح). نلاحظ أنه لو كانت البيانات معطاة لأقرب عشر (خانة عشرية واحدة) لكان نصف وحدة الدقة = 0.05 أما لو احتوت البيانات على خانتين عشريتين فإن وحدة الدقة تكون 0.01 ونصف وحدة الدقة تكون 0.005 وهكذا.

5- نعين الحد الأعلى الفعلي للفئة الأولى وذلك بإضافة طول الفئة إلى الحد الأدنى الفعلي لتلك الفئة.

إذن الحد الأعلى الفعلي للفئة الأولى هو 5 + 21.5 = 26.5

ثم نعين الحد الأعلى للفئة الأولى وهو يساوي الحد الأعلى الفعلي ناقصاً نصف وحدة الدقة ، ولذا فالحد الأعلى للفئة الأولى هو 26 = 0.5 - 26.5، وبهذا نكون قد حصلنا على الفئة الأولى وهي الفئة التي حدودها 22 ، 26 وحدودها الفعلية 21.5، 26.5.

6- نعين الحدود الدنيا والعليا لجميع الفئات الباقية وذلك بإضافة طول الفئة لكل حد. ثم نعين الحدود الدنيا الفعلية والعليا الفعلية وذلك بإضافة طول الفئة لكل حد فعلي.

7- نعين مراكز الفئات x_i ونجد مركز أي فئة بقسمة مجموع حديها على 2 ولذلك فمركز الفئة الأولى هو $24 = \frac{22+26}{2}$ ويمكن إيجاد مركز أي من الفئات

الأخرى بإضافة طول الفئة إلى مركز الفئة التي قبلها فيكون مركز الفئة الثانية 29 = 5+ 24 ومركز الفئة الثالثة 34 = 5 + 29 وهكذا.

8- نفرغ البيانات المعطاة لدينا على الفئات التي أنشأناها وذلك باستعمال خط عمودي لكل قراءة وخط مائل للقراءة الخامسة في كل فئة وذلك لتسهيل جمع التكرارات.

9- نجمع التكرارات المقابلة لكل فئة ونسجله في عمود التكرارات f_i ثم نجمع التكرارات لجميع الفئات ونقارنه بعدد البيانات فإذا كان عدد البيانات n فإنه يجب أن يكون: $\sum_{i=1}^{h} f_i = n$ حيث h عدد الفئات.

يظهر توضيح هذه الخطوات في الجدول (5)

الجدول (5)

التوزيع التكراري لعلامات ستين طالباً

حدود الفئة	الحدود الفعلية للفئة	مركز الفئة x_i	إفراغ البيانات	التكرار f_i
22 - 26	21.5 - 26.5	24		10
27 - 31	26.5 - 31.5	29		8
32 - 36	31.5 - 36.5	34		12
37 - 41	36.5 - 41.5	39		9
42 - 46	41.5 - 46.5	44		13
47 - 51	46.5 - 51.5	49		8
المجموع				60

وعند عرض التوزيع التكراري لا نكتب عمود إفراغ البيانات وفي بعض الأحيان لا نكتب عمود الحدود الفعلية بل نكتفي بعرض العمودين (1) و (5) أو العمودين (2) و (5) أو (3) و (5). ويمكننا تلخيص خطوات بناء التوزيع التكراري فيما يأتي:

- نعين عدد الفئات المتساوية.

- نعين المدى.

- نعين طول الفئة وذلك بقسمة المدى على عدد الفئات ثم التقريب إلى أعلى.

- نعين الحد الأدنى للفئة الأولى ثم نطرح منه نصف وحدة دقة لنعين الحد الأدنى الفعلي للفئة الأولى.

- نعين الحد الأعلى الفعلي للفئة الأولى وذلك بإضافة طول الفئة إلى الحد الأدنى الفعلي لتلك الفئة، ثم نطرح منه نصف وحدة دقة لنعين الحد الأعلى للفئة الأولى.

- نعين الحدود الدنيا والعليا والدنيا الفعلية والعليا الفعلية للفئات الباقية وذلك بإضافة طول الفئة لكل حد على التوالي.

- نعين مراكز الفئات وذلك بتعيين مركز الفئة الأولى ثم إضافة طول الفئة له لتعيين مركز الفئة الثانية وهكذا.

حـــدها

الأدنى + حدها الأعلى

مركز الفئة الأولى = ________________

2

- نفرغ البيانات على الفئات.

- نسجل مجموع تكرارات كل فئة أمامها في عمود التكرارات.

- نجمع التكرارات لجميع الفئات ونقارنه بعدد البيانات.

2 : 4 التوزيـــع التكـــراري النســـبي Relative Frequency Distribution

إن التكرار النسبي لكل فئة هو نسبة تكرار تلك الفئة إلى مجموع التكرارات، فإذا كان مجموع



التكرارات n وكان تكرار فئة معينة f فإن تكرارها النسبي هو $p = \frac{f}{n}$.

إن الجدول الذي يعطينا الفئات (أو مراكزها) مع تكراراتها النسبية يسمى التوزيع التكراري النسبي، وعليه فإن الجدول (6) يمثل التوزيع التكراري النسبي للمثال (7).

الجدول (6)

التوزيع التكراري النسبي لعلامات 60 طالباً.

حدود الفئة	التكرار النسبي
22 - 26	$\frac{10}{60} = 0.17$
27 - 31	$\frac{8}{60} = 0.13$
32- 36	$\frac{12}{60} = 0.20$
37 - 41	$\frac{9}{60} = 0.15$
42 - 46	$\frac{13}{60} = 0.22$
47 - 51	$\frac{8}{60} = 0.13$

ونلاحظ أن مجموع التكرارات النسبية يجب أن يساوي 1

أي أن $\sum p_i = 1$

5 : 2 التوزيع التكراري المتجمع Cumulative Frequency Distribution

نحتاج في كثير من الأحيان معرفة عدد المشاهدات التي تساوي قيمة معيّنة أو تكون أصغر منها. فعلى سبيل المثال، إذا حصل طالب على العلامة 80 في أحد الامتحانات فإنه يرغب في معرفة عدد الطلبة الحاصلين على العلامة 80 أو أقل في ذلك الامتحان.

وفي مثال (7)، ربما نحتاج الإجابة على أسئلة من نوع: كم طالباً حصل على العلامة 36 أو أقل؟

بالنظر إلى التوزيع التكراري في جدول (5) يكون عدد الطلاب= 30 =12+8+10

وهذا هو التكرار المتجمع للفئة الثالثة.

وبهذا فإن مجموع تكرارات جميع القيم التي تساوي الحد الأعلى الفعلي لفئة ما أو تقل عنه هو التكرار المتجمع لتلك الفئة.

إن الجدول الذي يعطينا الحدود الفعلية للفئات مع التكرارات المتجمعة المقابلة لها يسمى التوزيع التكراري المتجمع، ونبدأ دائماً بالحد الأدنى الفعلي للفئة الأولى ونعتبر تكراره المتجمع صفراً حيث لا يوجد بيانات تقل قيمتها عن ذلك الحد أو تساويه.

وعليه فإن الجدول (7) يمثل التوزيع التكراري المتجمع للمثال (7).

الجدول (7)

الحدود الفعلية للفئات	التكرار المتجمع
أقل من 21.5	0
أقل من 26.5	10
أقل من 31.5	18
أقل من 36.5	30
أقل من 41.5	39
أقل من 46.5	52
أقل من 51.5	60

أما إذا استعملنا التكرارات النسبية بدلا من التكرارات فإننا نحصل على التوزيع التكراري المتجمع النسبي، ونحصل على هذه النتيجة نفسها لو قسمنا التكرار المتجمع لكل فئة على مجموع التكرارات.

2 : 6 تمثيل التوزيعات التكرارية بيانيا

Graphical Presentation of Frequency Distributions

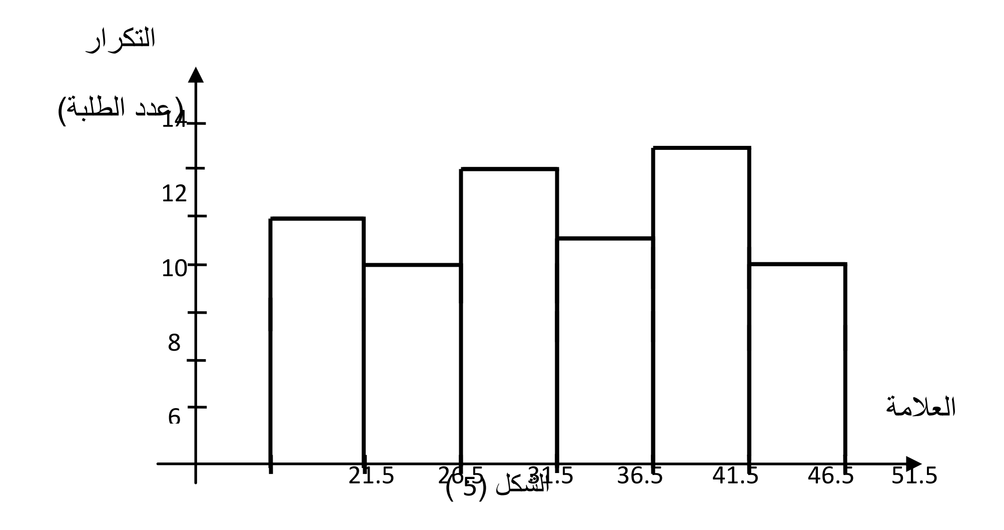
هناك ثلاث طرق لتمثيل التوزيعات التكرارية بيانياً

(1) المدرج التكراري Histogram

وهو عبارة عن تمثيل تكرار كل فئة من فئات التوزيع التكراري بمستطيل بحيث يتناسب تكرار كل فئة مع مساحة المستطيل المقام على تلك الفئة أي يكون ارتفاع

المستطيل = تكرار الفئة / طول الفئة

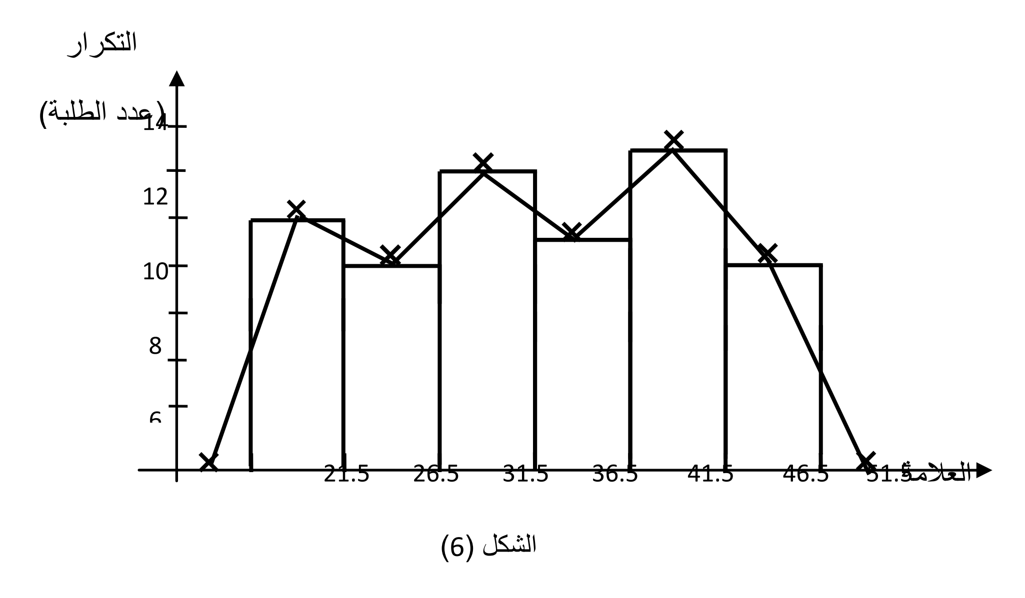
وإذا كان التوزيع التكراري ذا فئات متساوية فإننا نمثل كل فئة بمستطيل حدود قاعدته الحدود الفعلية لتلك الفئة وطول ارتفاعه يتناسب مع تكرارها. أي أننا نأخذ محورين متعامدين، نرصد على المحور الأفقي الحدود الفعلية لكل فئة من فئات التوزيع التكراري ونقيم على كل فئة مستطيلاً يتناسب طول ارتفاعه مع تكرار تلك الفئة كما هو موضح في الشكل (5) الذي يمثل المدرج التكراري المعطى في جدول (5).



الشكل (5)

(2) المضلع التكراري Frequency Polygon

المضلع التكراري هو مضلع مغلق نحصل عليه بتنصيف الأضلاع العلوية للمستطيلات في المدرج التكراري ثم بوصل هذه النقاط بعضها مع بعض، ولكي نغلق الخط المنكسر الذي حصلنا عليه نعتبر أن هناك فئتين متطرفتين واحدة إلى أقصى اليسار والثانية إلى أقصى اليمين وتكرار كل منهما صفر أي أن ارتفاع كل من المستطيلين المقامين على هاتين الفئتين صفر. نأخذ مركز كل من هاتين الفئتين ونغلق المضلع كما في الشكل (6) الذي يمثل المضلع التكراري للتوزيع التكراري في جدول (5).



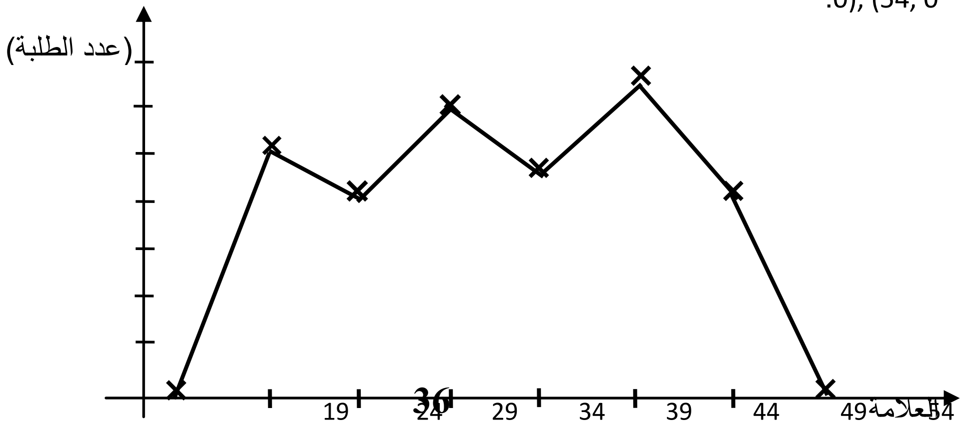
الشكل (6)

ويتضح من الشكل أن مجموع مساحات المستطيلات يساوي المساحة تحت المضلع التكراري. وهناك طريقة أخرى لرسم المضلع التكراري وذلك بأخذ محورين متعامدين، يمثل المحور الأفقي وحدات الفئات ويمثل المحور العمودي التكرارات.

نعتبر مركز كل فئة احداثياً أفقياً لنقطة ونعتبر تكرار هذه الفئة الإحداثي العمودي لتلك النقطة. نرصد جميع هذه النقاط ونوصل فيما بينها بخطوط مستقيمة.

نعين مركز فئة قبل الأولى مباشرة ونعتبر تكرارها صفراً ونعين مركز فئة بعد الفئة الأخيرة مباشرة ونعتبر تكرارها صفراً. نرصد هاتين النقطتين ونغلق بواسطتهما المضلع التكراري كما هو موضح بالشكل (7). أما النقاط فهي:

(49, 8) (44, 13) ,(39, 9) ,(34, 12) ,(29, 8) ,(24, 10). والنقطتان الإضافيتان هما : (19, 0), (54, 0).



الشكل (7)

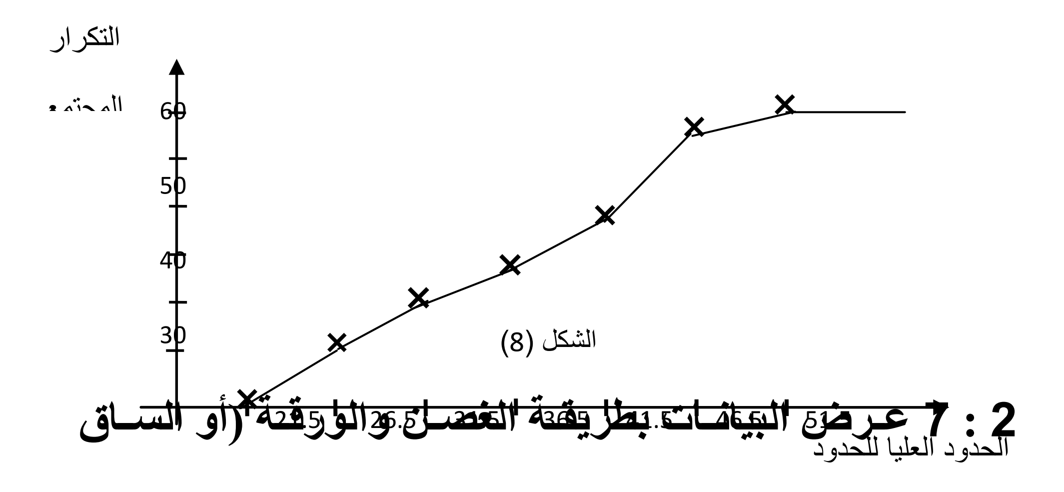
(3) المنحنى التكراري Frequency Curve

إذا مهدنا المضلع التكراري وجعلناه منحنىً بدلا من خطوط منكسرة فإننا نحصل على المنحنى التكراري. ويلاحظ أنه ينبغي عدم رسم المنحنى التكراري إلا إذا كانت الفئات كثيرة العدد، ذات طول صغير وكان عدد البيانات كبيراً وكانت هذه البيانات من النوع المتصل مثل الزمن والوزن.

وللمنحنى التكراري والمساحة المحصورة بينه وبين المحور الأفقي أهمية كبيرة في دراسة الإحصاء. وبنفس الطريقة التي مثّلنا بها التوزيع التكراري بيانياً نستطيع تمثيل التوزيع التكراري النسبي وذلك باستعمال التكرار النسبي على المحور العمودي بدلا من التكرار.

(4) المضلع التكراري المتجمع Frequency Ogive

نحصل على المضلع التكراري المتجمع برصد التكرار المتجمع لأي فئة مقابل الحد الأعلى الفعلي لها ثم وصل هذه النقاط فيما بينها بخطوط مستقيمة. أي أننا نأخذ محورين متعامدين ونمثل وحدات الفئات على المحور الأفقي والتكرار المتجمع على المحور العمودي ثم نرصد جميع النقاط التي احداثياتها الأفقية الحدود العليا للفئات واحداثياتها العمودية التكرارات المتجمعة المقابلة بها. ففي المثال (7)، ومن الجدول (7) نرصد النقاط: ,(41.5, 39) ,(36.5, 30) ,(31.5, 18) ,(26.5, 10) ,(21.5, 0)
(51.5, 60) ,(46.5, 52) ونحصل على المضلع التكراري المتجمع كما في الشكل (8).



الشكل (8)

2 : 7 عـرض البيانـات بطريقة الغصـن والورقـة (أو السـاق

والورقة)

Stem - and - Leaf Display

تعتبر طريقة الغصن والورقة لعرض البيانات طريقة أكثر كفاءة من المدرج التكراري وخاصة إذا كانت البيانات أعداداً ذات خانتين. ونحصل على عرض البيانات بطريقة الغصن والورقة بأن نعرض البيانات بصفوف حسب الخانة اليسرى لكل قيمة من تلك البيانات كما هو موضح في المثال التالي:

مثال (8) اعرض البيانات التالية بطريقة الغصن والورقة

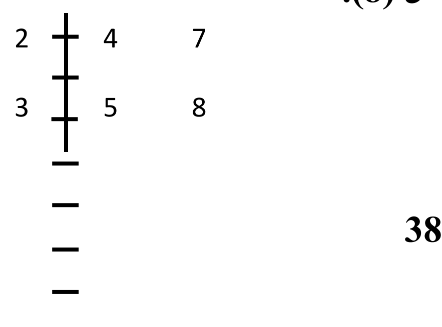
53	38	81	52	24
59	47	92	44	27
51	54	68	65	35
64	74	84	72	42
46	82	77	67	56

الحل:

1. ضع الأرقام من 2 إلى 9 مقابل خط عمودي. إن هذه الأرقام تقابل الخانة اليسرى لكل من البيانات وتسمى الغصون.

2. لكل قيمة من القيم، اكتب الرقم في الخانة الثانية (الورقة) لتلك القيمة مقابل القيمة على الخط العمودي (الغصن) الشكل 8 أ.

3. رتب القيم في كل صف حصلت عليه تصاعدياً. الشكل (8)ب يظهر العرض بطريقة الغصن والورقة للبيانات في المثال (8).



4	2	4	7	6		
5	6	2	4	3	9	1
6	5	7	8	4		
7	2	7	4			
8	1	4	2			
9	2					

الشكل 8 أ

2	4	7				
3	5	8				
4	2	4	6	7		
2	1	2	3	4	6	9
6	4	5	7	8		
7	2	4	7			
8	1	2	4			
9	2					

الشكل 8 ب

وبالنظر إلى العرض بطريقة الغصن والورقة فإننا نعتبر عمود الخانة الأولى إلى يسار الخط العمودي هو الغصن وقيم الخانة الثانية إلى يمين الخط العمودي هي الأوراق.وإذا ما نظرنا جانبياً لهذا العرض فإنه يظهر وكأنه مدرج تكراري طول فئته 10، ولكنه أكثر تعبيراً من المدرج لأن العرض بالغصن والورقة يوضح قيمة كل مفردة في البيانات، أي أننا

نستطيع معرفة قيمة كل مفردة في البيانات إذا كان لدينا العرض بالغصن والورقة وهذا لا يتأتي إذا أعطينا المدرج التكراري.

فمثلاً، إذا كان لدينا الغصن

4 0	3 4 7 9

فإن البيانات تكون 40, 43, 44, 47, 49

أي أننا استعدنا البيانات التي عرضت على ذلك الغصن.

وإذا كان لدينا الغصن

6.5 0	2 5 6 8

فهذا يعني أن البيانات هي

6.50, 6.52, 6.55, 6.56, 6.58

وكما أن من الممكن أن يكون الغصن ذا خانتين كما في الغصن السابق، فمن الممكن أن تحوي الأوراق خانتين، فمثلاً الغصن

0.3	02 15 24 31

هو عرض للبيانات

0.302, 0.315, 0.324, 0.331

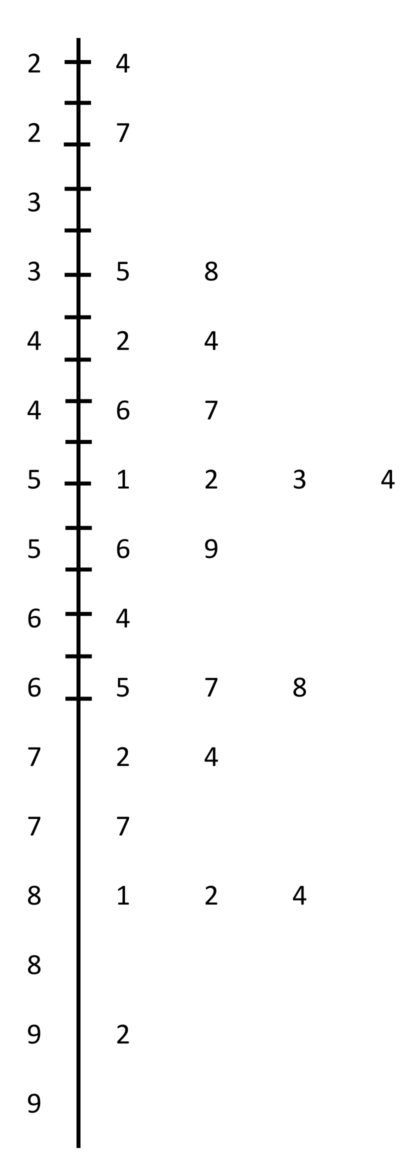
وإذا كانت البيانات كثيرة يمكننا أن نقسم كل غصن إلى غصنين كل منهما له الخانة الأولى (اليسرى) نفسها ونضع على الأول الأوراق التي خانتها الثانية من 0 إلى 4 ونضع على الغصن الثاني الأوراق التي خانتها الثانية من 5 إلى 9. وتسمى هذه الطريقة بطريقة الغصن المزدوج والورقة.

ويمكننا أن نضع الغصون تحت خط أفقي بدلاً من وضعها إلى يسار خط عمودي.

مثال (9): اعرض البيانات في مثال (8) بطريقة الغصن المزدوج والورقة.

الحل:نطبق الخطوات في مثال (8) مع الأخذ بعين الاعتبار أن الخانة اليسرى لكل غصن نكررها مرتين بحيث نضع البيانات التي فيها الخانة الثانية من 0 إلى 4 على الغصن الأول والبيانات التي خانتها الثانية من 5 إلى 9 نضعها على الغصن الثاني.

يظهر الحل في الشكل (9)



الشكل (9)

2 : 8 أشكال التوزيعات التكرارية

هناك ثلاث خواص يجب معرفتها عند وصف البيانات وهي الشكل والنزعة المركزية والتغّير. وسنبحث الآن شكل التوزيع التكراري:

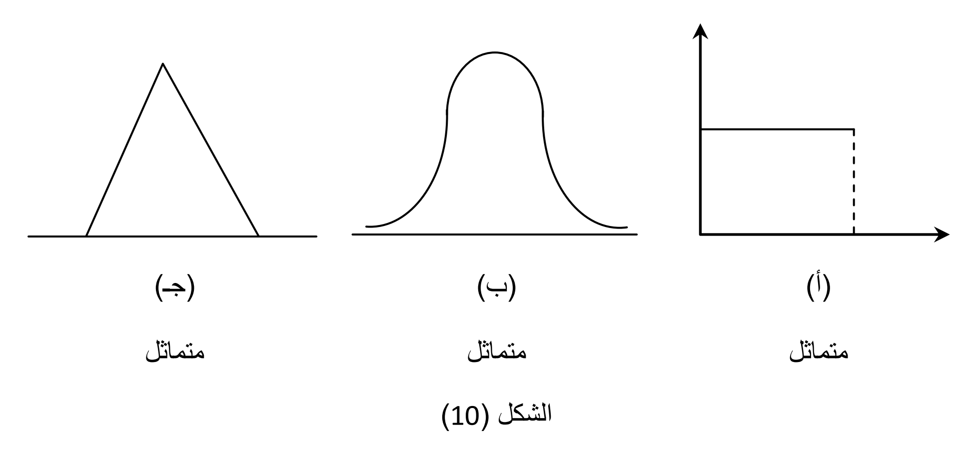
هناك أعداد لا نهائية من أشكال التوزيعات. وهناك (مقاييس) معاملات عددية يمكن استعمالها لمعرفة هذا التغير في الشكل بدقة وسنبحث بعضاً منها في هذا البند أما الآن فنكتفي بوصف خواص التوزيعات وصفاً إنشائياً مع أن هذا الوصف أقل وضوحاً من المقاييس العددية.

(1) إن أول تمييز في الشكل هو التمييز بين التوزيعات المتماثلة والتوزيعات غير المتماثلة.

ويعتبر التوزيع متماثلاً إذا أمكننا إقامة عمود على المحور الأفقي بحيث يقسم هذا العمود التوزيع إلى قسمين ينطبقان على بعضهما تمام الانطباق أي إذا أمكننا وضع مرآة عمودية على المحور الأفقي بحيث تقسم التوزيع إلى قسمين أحدهما صورة طبق الأصل للآخر. وبعبارة أخرى يكون التوزيع متماثلاً إذا أمكننا طيّه بحيث ينطبق أحد نصفيه على النصف الآخر تمام الانطباق والنقطة التي يكون الطي عندها أي النقطة التي يمكن إقامة عمود التماثل عليها تسمى نقطة التماثل.

أما في الحياة العملية فيوجد عدد قليل من التوزيعات المتماثلة ولكن يوجد كثير من التوزيعات التي تكون تقريباً متماثلة.

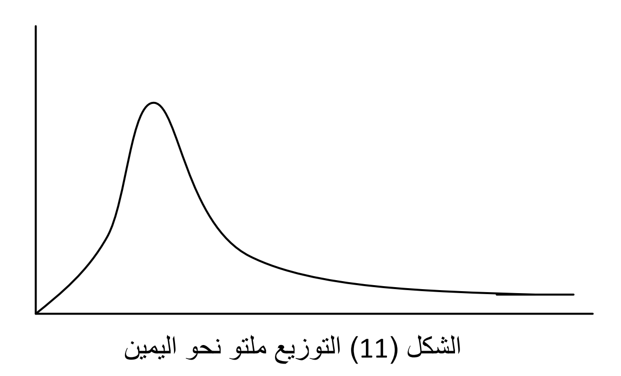
ويظهر من الشكل (10) أ، ب، جـ، د بعض التوزيعات المتماثلة.



الشكل (10)

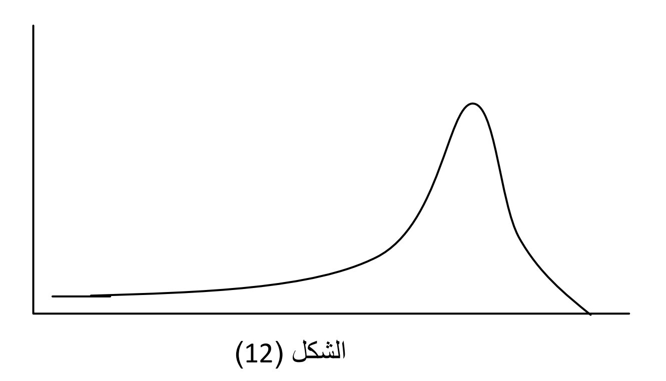
أما التوزيعات التي يكون عدم تماثلها واضحاً فتسمى توزيعات ملتوية. ويكون التوزيع ملتوياً إذا امتد أحد طرفيه إلى اليمين كثيراً أو امتد ذلك الطرف إلى اليسار كثيراً. وكذلك يكون التوزيع ملتوياً إذا كانت القيمة العليا فيه بعيدة عن المركز أي إذا كان عالياً من جهة ومنخفضاً من جهة ثانية.

إذا كان طرف التوزيع ممتداً إلى اليمين (أي في الاتجاه الموجب) نقول إن التوزيع ملتو نحو اليمين أو موجب الالتواء كما في الشكل (11).



الشكل (11) التوزيع ملتو نحو اليمين

أما إذا كان الطرف الطويل ممتداً نحو اليسار (أي في الاتجاه السالب) فنقول إن التوزيع ملتوٍ نحو اليسار أو سالب الالتواء كما في الشكل (12).



الشكل (12)

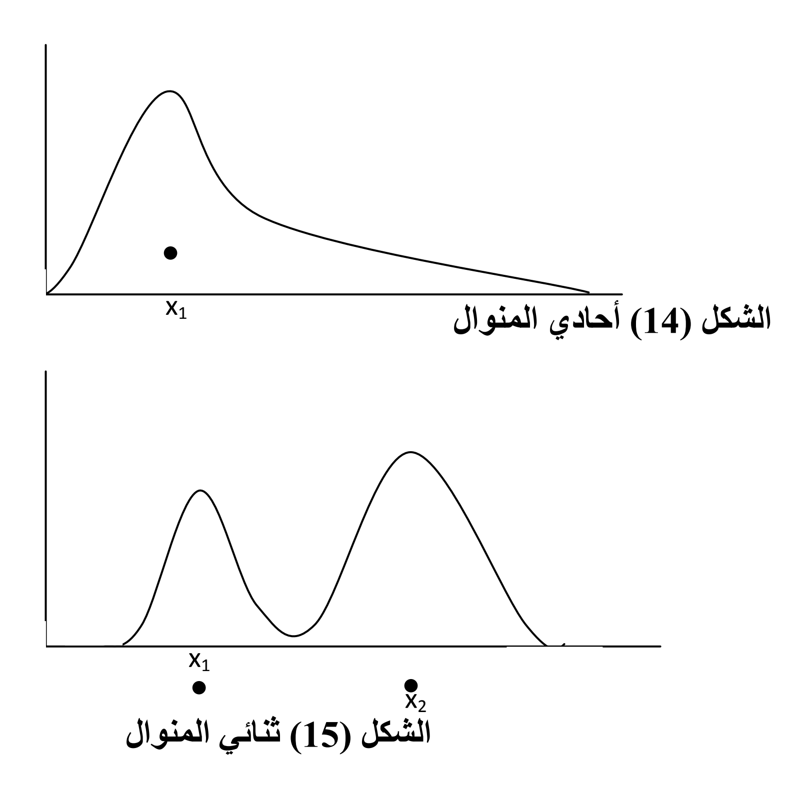
أما الشكل (13) فتوزيعه معتدل الالتواء.

الشكل (13)

(2) التمييز الثاني الذي نصف به التوزيعات التكرارية هو كونها ذات منوال واحد أو عدة منوالات وفي الحالة الأولى نسمى التوزيع أحادي المنوال وفي الثانية متعدد المنوالات وعندما يكون للتوزيع منوالان نقول إنه ثنائي المنوال، أما إذا كان التوزيع ذا ثلاثة منوالات فنقول إنه ثلاثي المنوال.

والمنوال هو القيمة التي يكون تكرارها أكبر من تكرار النقاط التي في جوارها أي أن المنوال هو النقطة على المحور الأفقي التي يقابلها قمة.

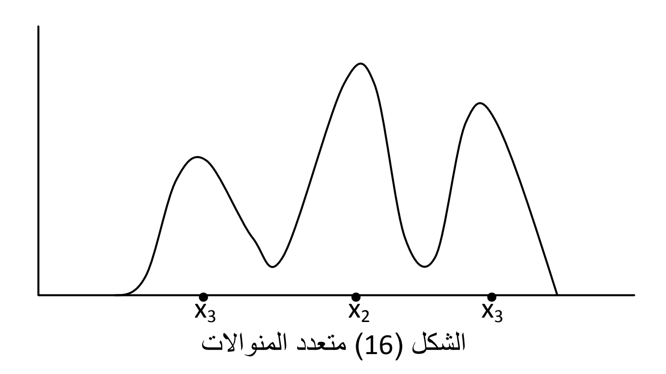
ففي الشكل (14) القيمة x_1 هي منوال لأنه يقابلها قمة وكذلك x_2 في الشكل (15) هما منوالان.



الشكل (14) أحادي المنوال

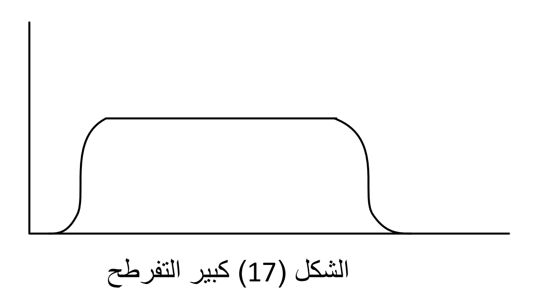
الشكل (15) ثنائي المنوال

أما في الشكل (16) فإن x_3, x_2, x_1 كلها منوالات لأنه يقابل كلا منها قمة أي أن التكرار على كل منها أكبر من التكرار على النقاط التي في جوارها.

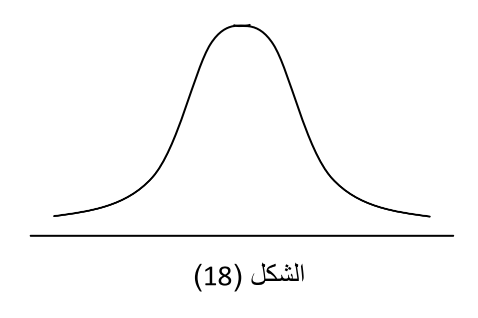


الشكل (16) متعدد المنوالات

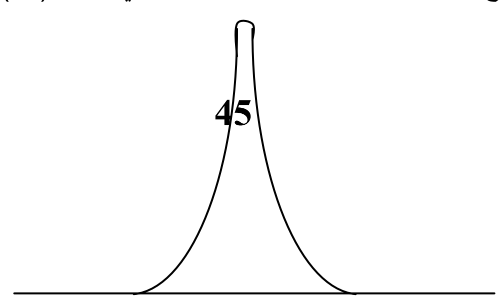
(3) التمييز الثالث الذي نصف به التوزيع التكراري هو تفرطح ذلك التوزيع حيث أن بعض التوزيعات تكون كبيرة التفرطح فتظهر على الشكل (17).


الشكل (17) كبير التفرطح

وبعضها يكون متوسط التفرطح كما يظهر في الشكل (18).


الشكل (18)

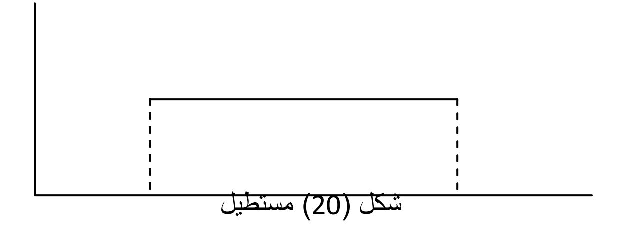
وبعضها يكون قليل التفرطح أو بعبارة أخرى يكون مدببا كما هو في الشكل (19).



الشكل (19) مدبب (قليل التفرطح)

(4) في كثير من الأحيان يوجد تسميات معيّنة لبعض التوزيعات التكرارية وهذه التسميات تصف لنا التوزيع وصفاً دقيقاً بدون ذكر صفات التماثل والالتواء والمنوال. وأكثر ما تتضح هذه التسميات في الأمثلة الآتية:

أ. توزيع متجانس (مستطيل) وشكله كما في الشكل (20).

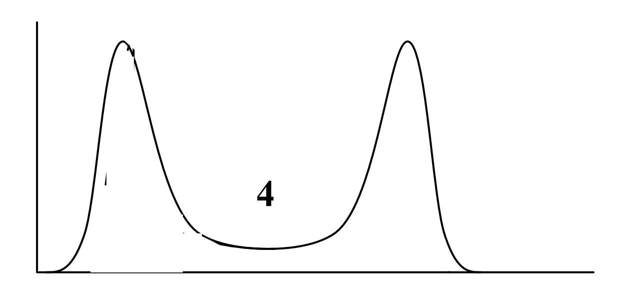


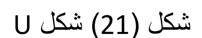
شكل (20) مستطيل

فعندما نقول إن هذا التوزيع متجانس فإننا نعلم أنه متماثل وأنه ليس له منوال واحد.

ب. توزيع U وهو كما في الشكل (21).

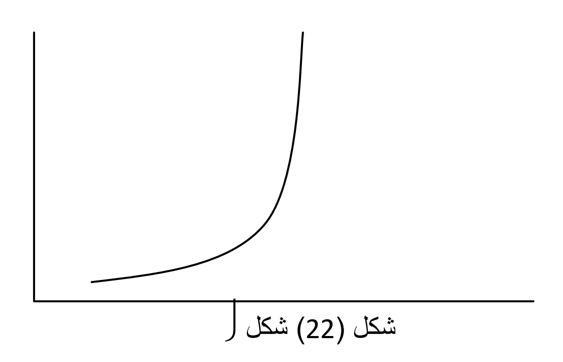
عندما يكون لدينا توزيع على شكل U فإننا نعلم أنه متماثل وأنه ثنائي المنوال أي أن له قمتين مرتفعتين من اليمين ومن الشمال. كما يظهر في الشكل (21).





شكل (21) شكل U

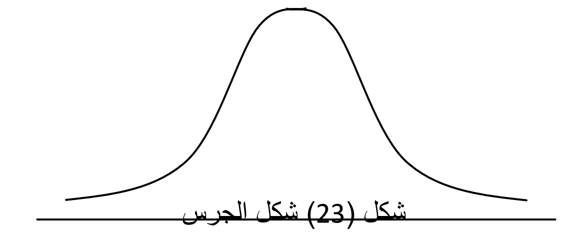
جـ- توزيع ل وهو كما في الشكل (22)



شكل (22) شكل ل

فـإذا كـان التوزيـع علـى شـكل ل والتـواؤه إلـى اليسـار فهـذا يعنـي أنـه أحادي المنوال وتقع القمة في أقصى اليمين أمـا إذا كـان التـواؤه إلـى اليمـين فهو أحادي المنوال وتقع القمة في أقصى اليسار.

د. توزيع شكل الجرس وهو كما في الشكل (23).



شكل (23) شكل الجرس

وهو توزيع متماثل وأحادي المنوال وتماثله يقع حول المنوال.

بعد التعرف على وصف مميزات التوزيعات التكرارية بالنظر إلى مضلعاتها التكرارية أو مدرجاتها التكرارية أو بالنظر إلى التوزيعات التكرارية نفسها ومقارنة التكرارات فيها، وبعد معرفة بعض التسميات لعدة أنواع من هذه التوزيعات بقي أن يمرّن الطالب نفسه على معرفة وصف أي توزيع تكراري وإعطاء فكرة عن شكله العام بمجرد النظر إلى ذلك التوزيع أو النظر إلى مضلعه التكراري.

تمارين

1-2. كان توزيع أعضاء هيئة التدريس في إحدى الجامعات الأهلية عام 1998/ 1997 على الكليات المختلفة كما يلي:

الكلية	العدد
العلوم	24
الآداب	18
الصيدلة	21
الهندسة	30
الحقوق	27

اعرض هذه البيانات بطريقة المستطيلات وبطريقة الدائرة.

2-2. كانت قياسات تركيز الأوزون (أجزاء لكل مائة مليون) في جو إحدى المدن كما يلي:

2.0	3.8	5.5	3.4
3.1	4.9	4.8	5.7
2.8	4.3	6.7	3.6
4.0	5.8	7.6	6.2
6.1	6.4	5.3	7.1

ضع هذه البيانات في توزيع تكراري ذي فئات متساوية بادئاً من الفئة 2.9 - 2.0 وأوجد التوزيع التكراري النسبي والتوزيع التكراري المتجمع.

3-2. أرسم المدرج التكراري والمضلع التكراري للتوزيع في التمرين (2-2).

4-2. اعرض البيانات في التمرين (2-2) بطريقة الغصن والورقة.

5-2. كانت علامات 36 طالباً في الامتحان النهائي في أحد المقررات كما يلي:

65	71	97	41	37	55
78	45	35	57	53	94
43	38	68	69	61	86
81	83	43	78	83	79
92	66	96	86	76	80
52	91	38	75	86	40

اعرض هذه البيانات بطريقة الغصن والورقة ثم بطريقة الغصن الثنائي والورقة.

أعط وصفاً تقريبياً لتوزيع البيانات بكلٍ من الطريقتين.

2-6. ضع العلامات في تمرين (2-5) في توزيع تكراري متساوي الفئات على أن تكون الفئة 59-50 إحدى فئات ذلك التوزيع.

كم عدد الفئات التي استعملتها وما طول الفئة وما هو التكرار النسبي للفئة 59-50 ؟

2-7. أرسم مضلع التكرار المتجمع للتوزيع في التمرين (2-6).

2-8. إذا كان لديك التوزيع التكراري التالي:

التكرار	حدود الفئات
1	8-19
2	20-31
5	32-43
8	44-55
11	56-67
14	68-79
6	80-91

أ. أرسم المضلع التكراري لهذا التوزيع.

ب. أعط وصفاً لهذا التوزيع.

جـ- أوجد التوزيع التكراري النسبي وأرسم مضلّعه.

2-9. عرضت مجموعة من البيانات بطريقة الغصن والورقة كما (الساق والورقة) في الشكل المقابل.

3	3					
4	2	6				
5	3	5	5			
6	2	4	6	8	9	9
7	4	4	7	8	9	
8	2	3	6	8		
9	4	5	7			

أ) اكتب البيانات.

ب) اعرض البيانات بطريقة الغصن الثنائي والورقة.

2-10. بلغت أعداد الطلبة الذين قبلوا في كلية العلوم حسب التخصصات في الفترة 99/ 98 - 96/ 95 كما في الجدول التالي:

التخصص	العام							
	95 /96		96 /97		97 /98		98 /99	
	ذكور	إناث	ذكور	إناث	ذكور	إناث	ذكور	إناث
رياضيات	50	30	45	40	52	44	55	45
فيزياء	35	35	40	42	45	40	46	44
كيمياء	60	40	55	50	60	55	65	60
علوم حياتية	50	40	60	45	55	50	50	60

إحصاء	30	20	25	30	35	30	30	30
حاسب إلكتروني	70	60	72	65	75	70	75	75

أ) اعرض مجموع الطلبة المقبولين في كلية العلوم حسب التخصصات في الفترة 99/ 98 - 96/ 95 بثلاث طرق.

ب) اعرض أعداد الطلبة المقبولين في الرياضيات حسب الجنس وسنة القبول بطريقة المستطيلات وطريقة الخط المنكسر.

ج) أجب عن فرع (ب) للطلبة المقبولين في الفيزياء.

د) اعرض مجموع الطلبة المقبولين في الحاسب الإلكتروني حسب السنوات بطريقة الدائرة.

هـ) أجب عن فرع (ب) لطلبة الكيمياء.

و) اعرض أعداد الطلبة المقبولين في العلوم الحياتية حسب الجنس وسنة القبول بطريقة المستطيلات.

ز) أجب عن فرع (و) لطلبة الإحصاء.

2-11. جاء في النشرة السكانية العدد الأول السنة الأولى 1995 أن معدل الوفاة للأطفال الرضع في الأردن كما يلي :

السنة	معدل وفيات الأطفال الرضع (لكل 1000 طفل ولد حيا)
1950	162
1960	151
1970	87
1980	70
1988	49
1990	45

اعرض هذه البيانات بطريقتين.

12-2. وجاء في النشرة السابقة أن توزيع السكان في المحافظات حسب الجنس لعام 1989 كما يلي :

المحافظة	ذكور	إناث
عمان	668230	628870
الزرقاء	231385	218515
اربد	387870	365530
المفرق	56160	52840
البلقاء	107900	106800
الكرك	67770	65030
الطفيلة	23510	22290
معان	57330	50970

أ. اعرض توزيع السكان حسب الجنس. والمحافظة بطريقة المستطيلات.

ب. اعرض توزيع السكان (ذكور وإناث والمجموع) حسب المحافظة بطريقة الخط المنكسر

13-2. كانت علامات الطلبة في أحد المساقات كما يلي :

81	57	92	36	50	44
70	50	42	78	33	71
91	68	71	85	69	75
89	50	54	51	34	73
73	76	69	68	53	63

77	72	87	50	87	69
61	72	39	65	79	94
50	64	63	65	69	46

أ. ضع هذه البيانات في توزيع ذي 5 فئات متساوية.

ب. ارسم المدرج التكراري والمضلع التكراري للتوزيع في (أ).

جـ. أوجد التوزيع التكراري النسبي والتوزيع التكراري المتجمع في (أ).

د. ضع البيانات في توزيع تكراري ذي فئات متساوية بادئا من الفئة 36-32.

هـ. ارسم المدرج التكراري والمضلع التكراري للتوزيع في (د).

14-2. كان أعضاء هيئة التدريس في احدى الجامعات من الذكور والإناث ينتمون إلى الدول التالية : الأردن ، العراق، سوريا، مصر، تركيا، الباكستان، الولايات المتحدة الأمريكية، وكانت رتب أعضاء هيئة التدريس هي : مدرس، أستاذ مساعد، أستاذ مشارك، أستاذ. صمم جدولا لعرض عدد أعضاء هيئة التدريس حسب الجنس والرتبة ودولة الجنسية.

الفصل الثالث
مقاييس النزعة المركزية والتشتت
Measures of Central Tendency and Dispersion

3 : 1 مقدمة

افترض أن لديك عدداً من المشاهدات لظاهرة معينة، ولتكن نتائج تجربة ما أو علامات الطلبة الناجحين في فحص شهادة الدراسة الثانوية في بلد ما في إحدى السنوات، والمطلوب منك أن تعطي تحليلا لهذه النتائج.

لا شك أن أول عمل تقوم به هو وضع هذه النتائج في جدول، لكنك سرعان ما تدرك كثرة مفردات ذلك الجدول، وصعوبة استنتاج أي فائدة منه. لذا فإنك تعمد إلى وضع هذه البيانات في توزيع تكراري غير مجمع، أي أنك ترتب البيانات ترتيبا تصاعديا وترصد تكرار كل علامة، من أدنى علامة إلى أعلاها.
لا تكون قد خسرت أية معلومة من البيانات، لكن إعطاء تحليل دقيق وسريع لها لا يزال صعبا لذلك عليك أن تضع هذه البيانات في توزيع تكراري ذي فئات متساوية وعددها معين ومعقول إنك بهذه الخطوة تكون قد كّثفت البيانات وجعلت طريقة عرضها أسهل وإن خسرت بعض المعلومات عنها.
فأنت الآن لا تعلم التكرارات التي تقابل كل قيمة داخل الفئة، لكنك تعلم عدد المفردات (التكرار) للفئة بكاملها. إنك تضحي بشيء من المعلومات في سبيل تسهيل عرض البيانات وسرعة الاستفادة منها.

حالما تضع البيانات في توزيع تكراري ذي فئات يصبح بالإمكان دراسة شكل التوزيع وطبيعته كالتماثل والالتواء وعدد المنوالات. لكن هذا لا يكفي، فأنت تود إيجاد مقاييس عددية تصف لك البيانات التي بين يديك. وأول المقاييس التي تفكر فيها مقاييس النزعة المركزية أو التوسط أو المعدلات. وهي عدة مقاييس عددية تعيّن موقع التوزيع، فقد يكون هناك توزيعات تكرارية متشابهة في طبيعتها وشكلها، لكنها تختلف في مواقعها. وفي مثل هذه الحال تكون معرفة المعدلات أو مقاييس النزعة المركزية ذات فائدة في دراسة الفرق بين هذه التوزيعات التكرارية.

على هذا الأساس يمكن القول إن المعدلات تعطيك قيمة تمثل العينة أو المجتمع الذي تدرسه، لكن بصورة غير دقيقة. ويتبلور هذا التمثيل إلى حد ما في إيجاد القيمة التي يتمركز حولها معظم المشاهدات. وهذه القيمة المتوسطة تمثل المجتمع أكثر من أية وحدة من مفرداته. وتتلخص الاستعمالات الرئيسة للمعدلات فيما يلي:

1- إنها تعطيك قيمة حالة أو ظاهرة معينة تمثل مجموعة معينة من الأفراد، وعادة ما تكون هذه القيمة مفيدة في الحياة العملية. فأنت مثلا تستطيع تحديد عدد العمال المطلوبين لإنجاز عمل معين إذا علمت معدل ما يستطيع العامل الواحد القيام به في اليوم.

2- إنها تعطيك قاعدة لمقارنة حالة معينة لمجموعة تحصيل مع نفس تلك الحالة لمجموعة ثانية. فمثلا لمقارنة تحصيل صف معين مع صف آخر، يمكنك استعمال معدل كل منهما كقاعدة للمقارنة.

3- إنها تسمح بتقدير حالة أو صفة لأفراد كثيرين إذا علمت قياسات عن جزء من المجموع الكلي لهؤلاء الأفراد، فمثلا إذا أردت تقدير حياة نوع من المصابيح الكهربائية التي ينتجها مصنع ما، فإنك تأخذ عددا من هذه المصابيح وتسجل حياة كل منها ثم تستعمل معدل الحياة للعينة المدروسة كتقدير لمعدل حياة ذلك النوع من المصابيح.

ويجب أن لا يغيب عن ذهنك أن المعدلات تعطي صورة "جزئية" عن المعلومات المحتواة في البيانات.

وإذا ما حقق المعدل (المتوسط) كل الصفات الآتية أو معظمها فإنه يعتبر مقبولا وهذه الصفات هي:

1- يجب أن يكون المتوسط معرّفا تعريفا دقيقا.

2- يجب أن يبنى على جميع أو معظم المشاهدات.

3- يجب أن يكون من السهل فهمه وتفسيره.

4- يمكن حسابه بسهولة وسرعة معقولتين.

5- يخضع للعمليات الجبرية بسهولة.

6- لا يتأثر كثيرا بالقيم المتطرفة أو الشاذة.

7-لا يتأثر كثيرا باختلاف العينات ذات الحجم الواحد من مجتمع واحد.

3 : 2 مقاييس النزعة المركزية

(1) الوسط الحسابي Arithematic Mean

تعريف (1)

إذا كان لدينا n من الأعداد (قيم المشاهدات)

$x_1, x_2, \dots, x_n$ فإن الوسط الحسابي لها هو حاصل قسمة مجموعها على عددها.

أي أن

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

حيث $\bar{x}$ يمثل الوسط الحسابي.

ويمكن استعمال الرمز$\sum$ (سيغما) الذي يعني جمع الحدود التي في داخله. وبذلك يصبح تعريف الوسط الحسابي.

$$\bar{x} = \sum_{i=1}^{n} x_i / n$$

مثال (1)

جد الوسط الحسابي للمشاهدات

12, 2, 22, 17, 0, 5, 15, 7

الحل: الوسط الحسابي يساوي

$$= 10\ \bar{x} = \frac{7+15+5+0+17+22+2+12}{8} = \frac{80}{8}$$

أما خواص رمز الجمع $\sum$ فهي:

تعني $\sum$ جمع الحدود التي في داخلها، وإذا كان المؤشر إلى أسفل يمين المتغير داخل $\sum$ يأخذ القيم من عدد صحيح إلى عدد صحيح آخر، فهذا يعني أن عليك أن تضع بدلا منه القيمة الأولى التي يأخذها المؤشر ثم تزيد واحدا تلو الآخر إلى أن تصل إلى العدد الأخير الموضوع فوق $\sum$.

فمثلا، إذا قلنا $\sum_{i=1}^{4} x_i$ فهذا يعني أن المؤشر هو i والقيم التي يأخذها هي i = 1

ثم i = 2 ثم i = 3 ثم i = 4 وبذلك يكون معنى $\sum_{i=1}^{4} x_i$ هو $x_1 + x_2 + x_3 + x_4$ في كثير من الأحيان لا تكون إشارة الجمع $\sum$ على متغيرات فيها مؤشر مثل x_i ولكن تكون عملية الجمع على اقتران في i مثل $\sum_{i=1}^{2}(i+5)^2$ أو $\sum_{i=1}^{2} i^3$ أو غيرها.

بالطبع فإن معنى $\sum$ يبقى "جمع الحدود في داخلها" وأنت تجد قيمة ذلك بأن تضع بدلا من i القيمة الدنيا لها، وهي القيمة الموجودة تحت $\sum$، ثم تضيف واحدا واحدا وتعوض بدلا من i إلى أن تصل إلى القيمة الموجودة في أعلى $\sum$.

فمثلا $\sum_{i=1}^{2}(i+2)^3$ يساوي $(2+2)^3 + (1+2)^3$

$= 91 = 3^3 + 4^3 = 27 + 64$

أما $\sum_{i=1}^{3}(i^2+5)$ فيكون معناها:

$$\sum_{i=1}^{3}(i^2+5) = (1^2+5) + (2^2+5) + (3^2+5)$$

$= (1 + 5) + (4 + 5) + (9 + 5)$

$= 6 + 9 + 14 = 29$

مثال (2) : أوجد قيمة $\sum_{i=1}^{4}(3i+2)$

الحل:

$$\sum_{i=1}^{4}(3i+2) = (3 \times 1 + 2) + (3 \times 2 + 2) + (3 \times 3 + 2) + (3 \times 4 + 2) =$$

$5 + 8 + 11 + 14 = 38$

ويمكن برهنة خواص $\sum$ التالية بسهولة وهي:

(1) إذا كان a ثابتا فإن $\sum_{i=1}^{n} a = n\,a$

وهذا يعني أنك إذا جمعت العدد الثابت a مرات عددها n فإن النتيجة تكون na مثل $\sum_{i=1}^{4} 5 = 4 \times 5$ $= 20$

(2) إذا كان a ثابتاً فإن $\sum ax_i = a\sum x_i$ أي أنك تستخرج الثابت خارج إشارة $\sum$.

$$\sum(x_i + y_i) = \sum x_i + \sum y_i \quad (3)$$

ملاحظة: لاحظ أنه إذا كان داخل الرمز $\sum$ حاصل ضرب متغيرين لهما نفس المؤشر فيجب إجراء عملية الضرب لكل قيمة من قيم i وبعدها تجمع حواصل الضرب، فمثلا لإيجاد $\sum_{i=1}^{2} x_i f_i$ خذ $i = 1$ وأوجد x_1f_1 ثم خذ $i = 2$ وأوجد x_2f_2 ثم اجمع هذين الحدين.

بذلك يكون $\sum_{i=1}^{2} x_i f_i = x_1f_1 + x_2f_2$ وستستفيد من هذا في تعريف الوسط الحسابي في حالة التوزيع التكراري أي البيانات المجمعة، حيث يكون معلوما لدينا حدود الفئات والتكرارات المقابلة لها وحيث تعتبر أن التكرار يتراكم على مركز الفئة. لذلك فإن أول عمل تقوم به هو إيجاد مراكز الفئات إذا لم تكن معطاة لنا . ويعرف الوسط الحسابي في حالة التوزيع التكراري كما يلي:

تعريف (2)

إذا كان لدينا توزيع تكراري عدد فئاته h وكانت مراكز الفئات (أو القيم) $x_2, x_1, \ldots, x_h$ والتكرارات المقابلة لها $f_2, f_1, \ldots, f_h$ فإن الوسط الحسابي يكون:

$$\bar{x} = \frac{x_1f_1 + x_2f_2 + \ldots + x_hf_h}{f_1 + f_2 + \ldots + f_h}$$

وباستعمال الرمز $\sum$ يكون :

$$\bar{x} = \frac{\sum_{i=1}^{h} x_i f_i}{n}$$

حيث $n = \sum_{i=1}^{h} f_i$ تمثل مجموع التكرارات.

مثال (3) :

أحسب الوسط الحسابي للتوزيع التكراري التالي:

مركز الفئة x_i	التكرار f_i	$x_i f_i$
5	7	35
10	10	100
15	4	60
20	12	240
25	9	225
30	3	90
المجموع	45	750

الحل: بما أن مراكز الفئات معطاة فأننا نحسب الوسط الحسابي مباشرة من التعريف:

$$\bar{x} = \frac{\sum x_i f_i}{n} = \frac{750}{45} = 16.67$$

مثال (4) :

جد الوسط الحسابي للتوزيع التكراري الآتي:

حدود الفئات	التكرار
12-16	10
17-21	13
22-26	07
27-31	05

الحل: جد مراكز الفئات أولا ثم أضرب مركز كل فئة في التكرار المقابل لها ورتب الحل كما في الجدول التالي

مركز الفئة x_i	التكرار f_i	$x_i f_i$
14	10	140
19	13	247
24	7	168
29	5	145
المجموع	35	700

ومن الجدول نجد أن الوسط الحسابي

$$\bar{x} = \frac{700}{35} = 20$$

(2) الوسيط The Median

(أ) الوسيط للبيانات الأولية (الخام)

في كثير من الأحيان ينصبّ اهتمام الدارس على معرفة القيمة التي يكون عدد الأفراد الحاصلين على قيمة أقل منها مساويا لعدد الحاصلين على قيمة أعلى منها، فمثلا: عندما تحصل على علامة معينة في

امتحان ما، هل تقع علامتك في النصف الأول من العلامات أم في النصف الثاني؟ في مثل هذه الحالة، عليك أولا أن تعرف تلك العلامة التي حصل نصف الطلبة على أقل منها ونصفهم الآخر على علامة أعلى منها. إذا عرفت هذه العلامة يكون بإمكانك الإجابة عن سؤالك السابق.

فمثلا إذا كانت العلامات التي حصل عليها طلبة الصف هي:

20, 6, 14, 10, 9, 7, 20, 18, 11, 10, 19, 17, 12, 13, 8

فما هي العلامة التي حصل نصف الطلبة على أقل منها والنصف الآخر على أعلى منها ؟ للإجابة عن هذا السؤال ترتب العلامات ترتيبا تصاعديا أي من الأصغر إلى الأكبر، فيكون لدينا:

6, 7, 8, 9, 10, 10, 11, 12, 13, 14, 17, 18, 19, 20, 20

من الواضح أن العلامة المطلوبة هي 12 فهناك سبع علامات أصغر منها وسبع أعلى منها. وتسمى العلامة 12 هذه الوسيط .

تعريف (3)

الوسيط لمجموعة من الأعداد المرتبة ترتيبا تصاعديا أو ترتيبا تنازليا هو العدد الأوسط منها إذا كان عددها فرديا، والوسط الحسابي للعددين الأوسطين في المجموعة إذا كان عددها زوجيا.

مثال (5) :

ما هو الوسيط للأعداد

19, 18, 2, 9, 8, 17, 5, 12, 11, 10

الحل : إن عدد البيانات زوجي، وبعد ترتيبها ترتيبا تصاعدياً تصبح:

2, 5, 8, 9, 10, 11, 12, 17, 18, 19 من الواضح أن العددين الأوسطين هما العددان الخامس والسادس وهما 11 ,10 إذن يكون الوسيط $\frac{10+11}{2}$ 10.5 = لاحظ إن هناك خمسة أعداد أقل من 10.5 وهي 10 ,9 ,8 ,5 ,2 وخمسة أعداد أكبر منها وهي 19 ,18 ,17 ,12 ,11.

تعريف (4)

إذا كان $X_1, X_2, \ldots, X_n$ مجموعة من الأعداد المرتبة تصاعديا (أو تنازليا) فإن الوسيط لهذه المجموعة هو العدد $X_{\frac{n}{2}+1}$ إذا كان n فردياً, وهو العدد $\frac{1}{2}(X_{\frac{n}{2}} + X_{\frac{n}{2}+1})$ إذا كان n زوجيا.

لاحظ أن $X_{\frac{n+1}{2}}$ هو العدد الذي رقمه هو $\frac{n+1}{2}$ وأن $X_{\frac{n}{2}}$ هو العدد الذي رقمه $\frac{n}{2}$ وذلك بعد ترتيب البيانات.

مثال (6) :

ما هو الوسيط للبيانات

18, 11, 17, 21, 25, 29, 35, 40, 5, 15, 9

الحل: بعد ترتيب البيانات حسب قيمها تصاعديا تصبح:

5, 9, 11, 15, 17, 18, 21, 25, 27, 35, 40

لاحظ أن عدد البيانات 11 أي فردي. إذن $\frac{11+1}{2}$ = 6 وبذلك يكون العدد السادس هو الوسيط وهو 18.

مثال (7) :

جد الوسيط للأعداد

7, 8, 10, 11, 12, 19, 23, 25

الحل: البيانات مرتبة ترتيبا تصاعدياً وعددها 8 (أي أنه زوجي) إذن فالوسيط هو:

الوسط الحسابي للعدديين الرابع والخامس، وبالتالي فإنه يساوي $\frac{11+12}{2} = 11\frac{1}{2}$

مثال (8) :

جد الوسيط للأعداد

2, 21, 19, 11, 17, 5, 14, 3, 6

الحل: رتب الأعداد ترتيبا تصاعديا فتصبح

2, 3, 5, 6, 11, 14, 17, 19, 21

والآن، لاحظ أن عدد البيانات 9 (أي أنه فردي). إذن الوسيط هو العدد الذي ترتيبه $\frac{9+1}{2}$ أي أنه العدد الذي ترتيبه الخامس، وبالتالي فالوسيط هو العدد 11.

(ب) الوسيط للتوزيع التكراري ذي الفئات

Median of Grouped Data (Frequency Distribution)

بعد أن تعرّفنا على طريقة إيجاد الوسيط من البيانات الأولية نحتاج إلى معرفة طريقة إيجاد الوسيط للتوزيع التكراري ذي الفئات، حيث يتم عرض كثير من البيانات بواسطة التوزيع التكراري.

ولما كان الوسيط هو القيمة التي يكون عدد التكرارات الأقل منها مساوياً لعدد التكرارات الأعلى منها ففي هذه الحالة نحتاج إلى معرفة التكرار المتجمع المقابل لأي فئة. نحصل على هذا التكرار بإيجاد مجموع التكرارات التي تساوي قيمها الحد الأعلى الفعلي لتلك الفئة أو تقل عنه. ونبدأ دائما بالحد الأدنى الفعلي للفئة الأولى، ونعتبر تكراره المتجمع صفرا، حيث لا توجد بيانات تقل قيمتها عن ذلك الحد أو تساويه.

مثال (9) :

جد التكرار المتجمع في التوزيع التكراري:

التكرار	حدود الفئات
5	10-13
11	14-17
8	18-21
6	22-25

الحل: نجد الحدود الفعلية للفئات ومن الواضح أن الحدود الفعلية للفئة الأولى هي 9.5-13.5 ومن الواضح أن عدد التكرارات التي قيمها أقل من 9.5 هو صفر. أما عدد البيانات التي قيمها أقل من 13.5 هو 5

استمر في جمع تكرار كل فئة لمجموع تكرارات الفئات التي تسبقها لتحصل على التكرار المتجمع كما في الجدول التالي:

الحدود الفعلية	التكرار المتجمع

أقل أو يساوي 9.5	0
أقل أو يساوي 13.5	5
أقل أو يساوي 17.5	16
أقل أو يساوي 21.5	24
أقل أو يساوي 25.5	30

تعريف (5)

إذا كان لديك توزيع تكراري مجموع تكراراته n فإن الفئة الوسيطية هي أول فئة يزيد تكرارها المتجمع عن $\frac{n}{2}$ أو يساويه.

مثال (10) :

ما هي الفئة الوسيطية في المثال (9)

الحل: n = 30، $\frac{n}{2}$ = 15

أنظر التوزيع التكراري المتجمع. ما هي أول فئة يزيد تكرارها المتجمع عن 15 أو يساويه؟ لاحظ أن العدد 16 هو الذي يزيد عن 15، لذلك فالفئة الوسيطية (13.5-17.5) ولإيجاد الوسيط للتوزيع التكراري. أفرض ما يلي:

مجموع التكرارات = n

طول الفئة الوسيطية = c

الحد الأدنى الفعلي للفئة الوسيطية = a

تكرار الفئة الوسيطية = f_m

التكرار المتجمع للفئة التي تسبق الوسيطية مباشرة = n_1 وبعبارة أخرى n_1 هو مجموع التكرارات التي تقل قيمها عن a.

تعريف (6)

الوسيط للتوزيع التكراري هو:

$$M = a + \frac{\frac{n}{2} - n_1}{f_m} \times c$$

حيث الرموز هي التي افترضناها سابقا

مثال (11) :

جد الوسيط للتوزيع التكراري للمصروف الشهري لستين طالبا المعطى في الجدول التالي:

التكرار	حدود الفئات
10	32-35
6	36-39
8	40-43
4	44-47
8	48-51
9	52-55
15	56-59

الحل: نكتب عمود التكرار المتجمع مقابل عمود الحدود الفعلية للفئات، فنحصل على الجدول:

التكرار المتجمع	التكرار	الحدود الفعلية للفئات
10	10	31.5-35.5
16	6	35.5-39.5
24	8	39.5-43.5

43.5-47.5	4	28
47.5-51.5	8	36
51.5-55.5	9	45
55.5-59.5	15	60

لاحظ أن عدد التكرارات $n = 60$ وطول الفئة $c = 4$. جد الفئة الوسيطية وهي أول فئة يكون تكرارها المتجمع $\frac{60}{2}$ أو أكثر. ومن الجدول تجد أن الفئة الوسيطية هي 47.5-51.5 إذن $a = 47.5$, $n_1 = 28$ وعليه $f_m = 8$

الوسيط يساوي:

$$M = 47.5 + \frac{\frac{60}{2} - 28}{8} \times 4$$

$$= 47.5 + \frac{8}{8} = 48.5$$

مثال (12) :

إذا كانت الأرباح اليومية لأربعين مكتبة كما يلي:

الحدود الفعلية (بالدينار)	التكرار
10.5-16.5	5
16.5-22.5	6
22.5-28.5	7
28.5-34.5	9

34.5-40.5	6
40.5-46.5	3
46.5-52.5	4

جد الوسيط.

الحل: نبني التوزيع التكراري المتجمع فيكون كما في الجدول.

الحدود الفعلية	التكرار	التكرار المتجمع
10.5-16.5	5	5
16.5-22.5	6	11
22.5-28.5	7	18
28.5-34.5	9	27
34.5-40.5	6	33
40.5-46.5	3	36
46.5-52.5	4	40

من الجدول: n = 40، طول الفئة c = 6 الفئة الوسيطية هي أول فئة يزيد تكرارها المتجمع عن أو يساوي $\frac{40}{2} = \frac{n}{2}$ 20 = وهي الفئة 28.5-34.5 وبالتالي فإن a = 28.5، $n_1 = 18$، $f_m = 9$.

الوسيط هو:

$$M = 28.5 + \frac{20-18}{9} \times 6 = 28.5 + \frac{12}{9} = 29.8$$

ويمكن إيجاد الوسيط بواسطة النسبة والتناسب في جدول التوزيع التكراري المتجمع مباشرة دون استعمال القانون ونوضّح ذلك بحل المثال (12).

نكتب التوزيع التكراري المتجمع كما يلي:

الحدود الفعلية	التكرار	التكرار المتجمع
10.5-16.5	5	5
16.5-22.5	6	11
22.5-28.5	7	18
M ←		→ 20
28.5-34.5	9	27
34.5-40.5	6	33
40.5-46.5	3	36
46.5-52.5	4	40

من الواضح أن الوسيط هو القيمة M التي يكون أقل منها أو يساويها $\frac{40}{2}$ 20 = تكراراً كما هو موضح بالسهم ↔

نجري التناسب التالي:

$$\frac{20-18}{27-18} = \frac{M-28.5}{34.5-28.5}$$

وبحل هذا التناسب نجد:

$$M = 28.5 + \frac{2}{9} \times 6 = 29.8$$

(3) المنوال The Mode

تعريف (7)

المنوال لمجموعة من القيم هو القيمة الأكثر تكرارا في توزيع تكراري للقيم أي أن المنوال يكون تلك القيمة التي يقابلها أكبر تكرار في جوارها.

مثال (13) :

إذا كانت الأجور اليومية التي يتقاضاها عدد من العمال بالدينار الأردني كما في الجدول التالي، جد المنوال.

الأجر اليومي	عدد العمال
4	4
5	5
6	10
7	18
8	9
9	8
10	3

الحل: **لاحظ أن القيمة 7 يقابلها أكبر تكرار في جوارها، ولذلك فإن المنوال = 7 دنانير.**

أما إذا لم يكن هناك قيم يقابلها تكرار أكبر من غيرها فلا يكون هناك منوال.

وفي بعض الأحيان يكون هناك أكثر من منوال، وتكرار كل منها أكبر من تكرار القيم في جواره. ومن الأمثلة التي يحدث فيها منوالان أن تحتوي البيانات على مجموعتين مختلفتين، لكن أفراد كل مجموعة لها قيم متقاربة من بعضها البعض. فمثلا، لو أخذت مجموعة من معلمي المدارس الإعدادية وأخرى من الأساتذة الجامعيين واعتبرت الرواتب للمجموعتين هي البيانات التي لديك. الواضح في تلك الحال أنك ستجد منوالين.

مثال (14) :

كانت رواتب عدد من معلمي المدارس الثانوية وعدد من أساتذة إحدى الجامعات ولأقرب دينار أردني كما يلي:

190, 910, 815, 190, 650, 650, 190, 175, 190, 185, 721, 650, 182

أوجد المنوال.

الحل: بالنظر إلى هذه الأرقام نجد منوال هذه الرواتب هو 190 حيث أنه تكرر أكثر من غيره من القيم التي في جوارها، وأن 650 تتكرر أكثر من غيرها أيضاً، ولذلك فإن 190 منوال و 650 منوال آخر.

وكما سبق وذكرنا في وصف أشكال التوزيعات التكرارية فإنه يمكن اعتبار المنوال هو القيمة التي يقابلها تكرار أكبر من تكرارات القيم في جوارها. فمثلا: إذا كان التوزيع الآتي يمثل علامات 34 طالبا في أحد الامتحانات (العلامة من 12) فإن العلامة 7 هي منوال لأن تكرارها أكبر من تكرار القيم التي في جوارها، وكذلك العلامة 11 منوال لنفس السبب. كما هو واضح في الجدول التالي.

العلامة x	التكرار f
5	2
6	3
7	9
8	4
9	5
11	8
12	3
المجموع	34

أما في التوزيعات التكرارية ذات الفئات فإننا نعطي التعاريف الآتية:

1- تسمى الفئة (أو الفئات) التي يقابلها أكبر تكرار الفئة المنوالية (الفئات المنوالية).

2- مركز الفئة المنوالية هو "المنوال التقريبي" وإذا كان هناك عدة فئات منوالية فإنه يكون لدينا عدة منوالات تقريبية.

مثال (15) : جد المنوال أو المنوالات التقريبية للتوزيع التالي:

التكرار	حدود الفئة
5	25-29
7	30-34
8	35-39
15	40-44
9	45-49
12	50-54
4	55-59

الحل: واضح أن الفئة 44-40 فئة منوالية، لأن تكرارها 15 وهو أكبر من تكرارات الفئات المجاورة لها، ولذلك فإن مركزها $\frac{40+44}{2}$ 42 = هو منوال تقريبي.

واضح أيضا أن الفئة 54-50 فئة منوالية، لأن تكرارها 12 أكبر من تكرارات الفئات المجاورة لها، ولذلك فمركزها $\frac{50+54}{2}$ = 52 منوال تقريبي أيضا.

إن المقاييس السابقة، الوسط الحسابي والوسيط والمنوال تعتبر كلها مقاييس موقع، حيث أنها تعين قيمة أو قيما تتمركز حولها البيانات بمعنى أو بآخر، وهي قيم لها نفس وحدات البيانات وعند رسمها بيانيا فإننا نرصدها على المحور الأفقي الذي نعين عليه حدود الفئات. ومن مقاييس الموقع أيضا المئينات والعشيرات والربيعات.

3 : 3 المئينات Percentiles

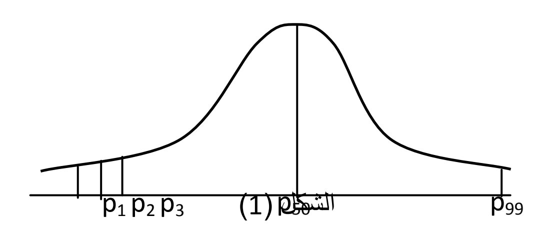
بعد دراستك للمضلع التكراري وبعض مقاييس النزعة المركزية للتوزيع التكراري تحتاج في كثير من الأحيان لمعرفة نسبة البيانات التي تقل عن قيمة معينة أو تساويها، أي أنك تريد الإجابة عن أسئلة من نوع :

ما نسبة عدد الطلبة الحاصلين على العلامة 95 أو أ قل في امتحان الشهادة الثانوية ؟ أو ما نسبة عدد الطلبة في صفك الذين يقضون أكثر من 40 ساعة في الدراسة البيتية أسبوعيا ؟ للإجابة عن هذه الأسئلة، ومثيلاتها، نقسم المحور الأفقي للمضلع التكراري إلى مائة قسم بحيث تكون المساحات تحت

المضلع على هذه الأقسام متساوية، أي تكون المساحة على كل قسم $\frac{1}{100}$ من المساحة الكلية تحت المضلع. إن النقاط التي تقسم المساحة تحت المضلع (أو المنحنى التكراري) إلى مائة قسم متساو في المساحة تسمى المئينات.

فالمئين الأول ورمزه p_1 هو القيمة التي يكون $\frac{1}{100}$ من البيانات أقل منها أو يساويها ويكون $\frac{99}{100}$ من البيانات أعلى منها أو يساويها، على فرض أن القيم مرتبة ترتيبا تصاعديا. إن تعريف المئينات الأخرى يكون بنفس الطريقة، فالمئين التسعون هو القيمة (النقطة على المحور الأفقي الذي هو محور القيم) التي يسبقها $\frac{90}{100}$ من البيانات ويليها $\frac{10}{100}$ منها على فرض أن البيانات مرتبة ترتيبها تصاعديا.

أما المئين الخمسون فهو القيمة التي تقسم البيانات إلى قسمين بحيث يكون نصف البيانات أقل من تلك القيمة (أو تساويها) ونصفها الآخر يكون أعلى من تلك القيمة (أو تساويها) الشكل (7) يوضح مفهوم المئينات، حيث المساحة تحت المنحنى على كل جزء تساوي $\frac{1}{100}$.



تعريف (8) :

إذا رتبت البيانات ترتيبا تصاعديا فإن المئين k (نعبر عنه بالرمز P_k) هو القيمة التي يكون أقل منها (أو يساويها) k بالمائة من البيانات وأعلى منها أو يساويها، (k - 100) بالمائة منها.

والسؤال الآن كيف تجد المئين k؟ أي كيف تجد P_k من التوزيع التكراري ذي الفئات، حيث k تأخذ القيم 1 ,2 , ... ,99 ؟

إن طريقة إيجاد P_k هي نفس طريقة إيجاد الوسيط، وذلك بأن تجد التوزيع التكراري المتجمع ثم تجد فئة المئين k ، وهي أول فئة يزيد تكرارها المتجمع على (أو يساوي) $\frac{k}{100}\times n$ حيث n هو مجموع التكرارات كلها.

عين الحد الأدنى لفئة المئين k وعبر عنه بالرمز a ثم استعمل القانون :

$$P_k = a + \frac{\frac{k}{100}\times n - n_1}{f}\times c$$

حيث :

n_1 = التكرار المتجمع للفئة التي تسبق فئة المئين k مباشرة.

f = تكرار فئة المئين k.

c = طول الفئة (طول فئة المئين k إذا لم يكن التوزيع ذا فئات متساوية).

لاحظ أنك حالما تعين فئة المئين k فإنك تحدد قيم جميع الرموز في الجهة اليمنى من المعادلة، وبتعويض هذه القيم تجد P_k.

مثال (16) :

لديك التوزيع التكراري في الجدول التالي الذي يمثل علامات 70 طالبا.

(3) التكرار المتجمع	(2) التكرار	(1) الحدود الفعلية للفئات
2	2	21.5-26.5
11	9	26.5-31.5
17	6	31.5-36.5
28	11	36.5-41.5
41	13	41.5-46.5
57	16	46.5-51.5
64	7	51.5-56.5
70	6	56.5-61.5

أوجد المئين الثمانين P_{80}.

الحل : أوجد التكرار المتجمع وضعه في العمود (3) ولاحظ أن n = 70.

لاحظ أن المئين الثمانين يسبقه ثمانون بالمائة من التكرارات، أي أن مجموع التكرارات التي تسبقه هي $\frac{80}{100} \times 70 = 56$.

عين فئة المئين 80 أي أول الفئة التي يزيد تكرارها المتجمع على 56 أو يساويه. من الواضح أن هذه الفئة هي : 46.5-51.5 وبالتالي فإن الحد الأدنى الفعلي لفئة المئين 80 هو a = 46.5.

ولاحظ أن التكرار المتجمع فئة تسبق فئة المئين 80 مباشرة هو $n_1 = 41$ وتكرار فئة المئين الثمانين هو f = 16.

وبالتعويض في القانون $P_k = a + \frac{\frac{k}{100} \times n - n_1}{f} \times c$

نجد : $P_{80} = 46.5 + \frac{56-41}{16} \times 5$

$$= 46.5 + \frac{15}{16} \times 5$$

$$= 51.2$$

وكما عرفنا المئينات نعرف العشيرات Deciles وهي القيم التي تقسم البيانات المرتبة تصاعديا إلى عشرة أقسام متساوية هذه القيم هي $D_1, D_2, \dots , D_9$ وبمقارنة العشيرات بالمئينات نجد أن $D_1 = P_{10}$, $D_2 = P_{20}$, ... , $D_9 = P_{90}$ ولذلك يمكننا استعمال المئينات للإجابة عن أي سؤال يتعلق بالعشيرات، ففي المثال السابق وجدنا P_{80} أي أننا وجدنا D_8.

إن بعض المئينات لها أهمية خاصة وهذه هي P_{25}, P_{50} , P_{75} حيث أن P_{25} يسمى الربيع الأول Q_1، P_{50} هو الربيع الثاني Q_2 أي الوسيط، P_{75} هو الربيع الثالث Q_3 وزيادة في الإيضاح نورد التعريف التالي :

تعريف (9) :

الربيعات Quartiles في البيانات المرتبة تصاعديا هي :

الربيع الأول Q_1 هو القيمة التي يسبقها ربع البيانات ويليها ثلاثة أرباع البيانات.

الربيع الثاني Q_2 وهو الوسيط، أي القيمة التي يسبقها نصف البيانات ويليها نصفها.

الربيع الثالث Q_3 وهو القيمة التي يسبقها ثلاثة أرباع البيانات ويليها ربعها.

ويتم حساب الربيع الأول Q_1 والربيع الثالث Q_3 بنفس الطريقة التي تم فيها حساب المئينات أو حساب الوسيط (الربيع الثاني Q_2) وذلك باستعمال الرموز والتعاريف التالية:

فئة الربيع الأول هي أول فئة يزيد تكرارها المتجمع عن $\frac{n}{4}$ من البيانات حيث n هو مجموع التكرارات.

طول فئة الربيع الأول = c

الحد الأدنى الفعلي لفئة الربيع الأول = a_1

تكرار فئة الربيع الأول = f_1

التكرار المتجمع للفئة التي تسبق فئة الربيع الأول = n_1

وبعبارة أخرى n_1 هو مجموع التكرارات قبل a_1

تعريف (10)

الربيع الأول هو $Q_1 = a_1 + \dfrac{\frac{n}{4} - n_1}{f_1} \times c$

وبنفس الطريقة: فئة الربيع الثالث هي أول فئة يزيد تكرارها المتجمع عن $\dfrac{3n}{4}$، الربيع الثالث هو:

$$Q_3 = a_3 + \dfrac{\frac{3n}{4} - n_3}{f_3} \times c$$

حيث الرموز كمثيلاتها السابقة، أي:

الحد الأدنى الفعلي لفئة الربيع الثالث = a_3

تكرار فئة الربيع الثالث = f_3

التكرار المتجمع للفئة التي تسبق فئة الربيع الثالث = n_3

مثال (17)

الجدول التالي يمثل التوزيع التكراري للزمن (لأقرب دقيقة) الذي استغرقه 50 طالبا للإجابة عن أسئلة امتحان شهري.

حدود الفئات (الزمن)(1)	التكرار (2)	التكرار المتجمع (3)
32-36	3	3
37-41	8	11
42-46	9	20
47-51	8	28
52-56	17	45

57-61	5	50

جد الوسيط (Q_2) والربيع الأول Q_1 والربيع الثالث Q_3.

الحل: أكتب التكرار المتجمع في العمود الثالث.

الفئة الوسيطية هي أول فئة يزيد تكرارها المتجمع عن $\frac{50}{2}$ أي 25 وبالتالي فهي الفئة 51-47 وحدها الأدنى الفعلي هو 46.5

بتطبيق تعريف الوسيط تجد

$$M = Q_2 = 46.5 + \frac{25-20}{8} \times 5$$

لأن تكرار الفئة الوسيطية هو 8 وطول تلك الفئة 5

إذا: $M = 46.5 + 3.1 = 49.6$

لحساب الربيع الأول، لاحظ أن فئة الربيع الأول هي أول فئة يزيد تكرارها المتجمع عن $\frac{50}{4}$ أي 12.5 وبالتالي فهي الفئة 46-42 [وحدها الأدنى الفعلي هو 41.5]

مجموع التكرارات قبل الحد الأدنى الفعلي هو $n_1 = 11$

طول الفئة $c = 5$

تكرار فئة الربيع الأول هو $f_1 = 9$

بتطبيق تعريف الربيع الأول

$$Q_1 = 41.5 + \frac{12.5-11}{9} \times 5$$

$= 42.3$

ولحساب الربيع الثالث، لاحظ أن فئة الربيع الثالث هي أول فئة يزيد تكرارها المتجمع عن $\frac{3}{4} \times 50$

أي 37.5 وبالتالي فهي الفئة 56-52 [وحدها الأدنى الفعلي

هو 51.5]

مجموع التكرارات قبل الحد الأدنى الفعلي هو $n_3 = 28$

طول الفئة $c = 5$

تكرار فئة الربيع الثالث هو $f_3 = 17$

بتطبيق تعريف الربيع الثالث:

$$Q_3 = 51.5 + \frac{37.5 - 28}{17} \times 5$$

$$= 51.5 + 2.8 = 54.3$$

لاحظ كذلك أنه يمكن حساب Q_1، Q_2، Q_3 عن طريق النسبة والتناسب كما حسبنا الوسيط سابقاً.

أما حساب المئينات (وبالتالي الربيعات والعشيرات) من البيانات الأولية، غير المجمعة في توزيع تكراري، فيكون بإتباع الخطوات التالية :

(1) نرتب البيانات ترتيبا تصاعديا ونحدد عددها وليكن n.

(2) لحساب المئين k أي لإيجاد P_k، نجد قيمة nk فإذا كان nk كسرا نقربه إلى أعلى (أي إلى العدد الصحيح الذي يأتي بعد nk مباشرة) ونأخذ المفردة التي رقمها ذلك العدد الصحيح، أما إذا كان nk عددا صحيحا فنأخذ القيمتين ذاتي الرقمين nk , nk+1 ونأخذ الوسط الحسابي لهما.

مثال (18) :

لديك البيانات التالية:

5, 8, 6, 15, 19, 4, 7, 11, 22, 21, 24, 13, 27, 19, 25

أوجد المئين السبعين، وأوجد Q_1.

الحل : نرتب البيانات تصاعديا، لاحظ ان عددها n =15.

4, 5, 6, 7, 8, 11, 13, 15, 19 , 19 , 21, 22, 24, 25, 27

المئين السبعون P_{70} ، إذن k = 0.7. ويكون nk = 15 x 0.7 = 10.5

إذن P_{70} هو القيمة التي ترتيبها 11 (تقريب 10.5 إلى أعلى هو 11) أي $P_{70} = 21$.

لإيجاد Q_1 ، نعلم أن $Q_1 = P_{25}$ ، إذن $k = 0.25$، $nk = 15 \times 0.25 = 3.75 \approx 4$ إذن Q_1 هو القيمة التي ترتيبها 4 أي $Q_1 = 7$.

3 : 4 مقارنة بين صفات الوسط الحسابي والوسيط والمنوال

Comparison Between The Measures of Central Tendency

إن أكثر مقاييس النزعة المركزية استعمالا في الإحصاء هو الوسط الحسابي. فهو سهل الحساب، سهل التعريف. كما يأخذ جميع قيم المتغير تحت البحث بعين الاعتبار.

وبما أن الوسط الحسابي يعرّف بمجموع القيم مقسوما على عددها فإنك تستطيع معرفة مجموع القيم إذا علمت الوسط الحسابي وعدد التكرارات، وهناك خاصية مهمة للوسط الحسابي وهي أن مجموع انحرافات البيانات عن وسطها الحسابي يساوي صفراً، أي أن:

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0 \text{ ؛ } \sum_{i=1}^{h}(x_i - \bar{x})f_i = 0$$

وبالنظر إلى المعادلة السابقة والمدرج التكراري للبيانات فإن الوسط الحسابي هو نقطة اتزان المدرج التكراري. وبما أن الوسط الحسابي هو نقطة اتزان التوزيع، فإنك لو أضفت أي عدد من القيم المساوية للوسط الحسابي إلى البيانات فإن هذه الإضافة لا تؤثر عليه، أما إذا أضفت مفردات تختلف قيمها عن قيمة الوسط الحسابي فإن قيمته تتغير.

فمثلا: إذا كانت علامات 12 طالبا في أحد الامتحانات كما يلي:

30, 28, 26, 19, 30, 17, 29, 20, 30, 21, 23, 27

فإن الوسط الحسابي للعلامات يساوي 25 فإذا أضفت العلامات 25 ,25 ,25 ,25, فإن الوسط الحسابي للطلبة جميعهم (أصبحوا 15 طالبا) يبقى 25 ولكن إذا أضفت طالبا واحدا إلى المجموعة الأصلية من الطلبة (الاثنى عشر طالبا) وكانت علامته 20 مثلا فإن الوسط الحسابي للمجموعة الكلية (13 طالبا) يتغير ويصبح في هذه الحالة:

$$\frac{27+23+21+30+20+20+17+30+19+20}{13}$$

$$= 24\ \frac{8}{13} = 24.62$$

أما أهم عيوب الوسط الحسابي فهي تأثره الشديد بالقيم المتطرفة، فلو قلت إن مجموعة من الأفراد تبرعوا لعمل خيري بمبالغ بسيطة (أقل من 10 دنانير من كل منهم) ثم تبرع اثنان بمبلغين كبيرين (100 من كل منهما) وحسبت الوسط الحسابي للتبرعات لوجدت أن معدلها كبير، وهذا لا يمثل ما دفعه معظم الأفراد.

أما الوسيط فهو سهل التعريف، سهل الحساب، لا يتأثر بالقيم الشاذة ولا يعتمد على جميع القيم دائما. إن تغير قيمة من القيم قد يؤثر فيه وقد لا يؤثر. وهو يستعمل خاصة في البيانات التي يعرف ترتيبها ولا تعرف قيمها، وكذلك في البيانات الناقصة، أي إذا كان في البيانات مفردات سقطت قيمها ولكن عرف ترتيبها. وأهم الأمثلة على ذلك، تلك البيانات التي تقيس الزمن الذي ينجز فيه الطلاب عملا معينا. فقد ينتهي الوقت المقرر دون أن يتمكن عدد من الطلبة من إنجاز ذلك العمل، ولذلك لا تكون هناك قيم لزمن إنجاز العمل عند هؤلاء الطلاب، لكنك تعرف أنهم يستغرقون وقتا أطول من باقي الطلبة أي أنك تعرف ترتيبهم، وبالتالي فأنت تستطيع إيجاد الوسيط لمثل هذه التجربة إذا كان عدد الذين لم ينجزوا العمل أقل من نصف المجموعة. ولا تستطيع إيجاد الوسط الحسابي.

إذا أخذت عينة من مجتمع ما، ثم عيّنة من نفس المجتمع فإنك تجد تقاربا بين الوسطين الحسابيين لهاتين العينتين أكثر من التقارب بين وسيطيهما أي أن الوسط الحسابي بصورة عامة أكثر ثبوتا من الوسيط. فالوسط الحسابي لا يتأثر كثيرا من عيّنة لأخرى، أما الوسيط فيتأثر. فإذا أردت تقدير معدل المجتمع باستعمال أحد مقاييس النزعة المركزية فإن الوسط الحسابي أكثر ثبوتا وقبولا للاعتماد عليه.

وهكذا فمن الممكن أن يكون بعد وسط العينة الحسابي عن الوسط الحسابي للمجتمع أصغر من بعد وسيط العينة عن وسيط المجتمع ولكن ليس هذا مؤكداً في جميع الحالات.

أما المنوال فهو أقل الثلاثة استعمالا، وهو في حال البيانات القليلة العدد عديم الفائدة تقريبا، هذا إن وجد أصلا.

أما في البيانات الكبيرة العدد فللمنوال معنى معقول، وميزته أنه لا يحتاج إلى حسابات إلا في حالة التوزيعات التكرارية ذات الفئات، كما أنه يتأثر كثيرا بطريقة اختيار الفئات، فمن الممكن أن توجد فئتان منواليتان غير متقاربتين، بصدفة اختيار الفئات فقط.

هذا من جهة ومن جهة ثانية فإن المنوال لا يتأثر كثيرا بالقيم الشاذة، بل أنه لا يتأثر في بعض الأحيان، حتى لو تغيّرت قيم بعض مفردات البيانات. هذا ويفضّل استعمال الوسط الحسابي في الحالات التالية:

أ- إذا كان التوزيع متماثلا تقريبا.

ب- إذا كان اهتمامك منصبا على القيمة العددية لجميع البيانات أي المجموع الكلي، بدلا من الاهتمام بقيمة نموذجية.

أما الوسيط فيفضل استعماله:

أ- إذا أردت إيجاد قيمة ممثلة، أي قيمة نموذجية بدلاً من اهتمامك بالمجموع الكلي، وإذا كان التوزيع ملتوياً.

ب- إذا فقدت لديك بعض القيم (وعرف ترتيبها) حيث لا يمكن حساب الوسط الحسابي مباشرة.

أما المنوال فلا يفضل استعماله، إلا إذا كان التوزيع متعدد المنوالات.

3 : 5 الوسط الحسابي المرجح لأوساط حسابية :

إذا كان لدينا مجموعات ذات أعداد مختلفة من البيانات وعلم الوسط الحسابي لكل مجموعة فكيف نحصل على الوسط الحسابي للمجموعات كلها إذا دمجت معا ؟

القاعدة التالية تعطينا الإجابة عند دمج مجموعتين معا.

قاعدة (1) :

إذا كان لدينا مجموعة ذات n_1 من القيم ووسطها الحسابي $\bar{x}_1$ ، ومجموعة ثانية ذات n_2 من القيم ووسطها الحسابي $\bar{x}_2$ ، فإن الوسط الحسابي للمجموعة ذات n_2+n_1 من القيم الناتجة من دمج المجموعتين يكون $\bar{x} = \dfrac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$ وهذا هو الوسط الحسابي المرجح (الموزون) لأوساط حسابية. وهذه القاعدة صحيحة لو دمجنا أي عدد محدود من المجموعات مع بعضها البعض.

مثال (19) : لو دمجت مجموعتان أ ، ب وكان عدد البيانات في أ يساوي 20، ووسطها الحسابي 45، وعدد البيانات في ب يساوي 30 ووسطها الحسابي 50 فإن الوسط الحسابي للمجموعة الناتجة من دمج المجموعتين يساوي :

$$\bar{x} = \frac{20 \times 45 + 30 \times 50}{20 + 30}$$

$$= \frac{2400}{50} = 48$$

لاحظ أن الوسط المرجح يقع بين وسطي المجموعتين.

ويمكن تعميم قاعدة (1) لإيجاد الوسط الحسابي المرجح لعدة مجموعات بعد دمجها.

قاعدة (2) :

إذا كان لدينا مجموعة من البيانات عددها n_1 ووسطها الحسابي $\bar{x}_1$ ومجموعة ثانية عدد أفرادها n_2 ووسطها الحسابي $\bar{x}_2$ ومجموعة ثالثة عدد أفرادها n_3 ووسطها الحسابي $\bar{x}_3$ ..... ومجموعة رقمها k وعدد أفرادها n_k ووسطها الحسابي $\bar{x}_k$ فإن الوسط الحسابي للمجموعة الناتجة من دمج المجموعات كلها هو :

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3 + ... + n_k\bar{x}_k}{n_1 + n_2 + n_3 + ... + n_k}$$

3 : 6 مقاييس التشتت Measures of Dispersion

لقد درسنا في الفصول السابقة طرق عرض البيانات الإحصائية وبناء التوزيعات التكرارية وعرضها بيانيا كما أوردنا وصف أشكالها وبعض خواصها. ثم درسنا مقاييس النزعة المركزية التي تصف هذه التوزيعات عدديا وهي تعتبر مقاييس موقع، أي أن قيمها تكون على المحور الأفقي الذي يمثل قيم البيانات.

إن هذه المقاييس غير كافية لتحديد صفات التوزيعات التكرارية والبيانات الإحصائية. فقد يكون لدينا ظاهرتان متساويتان في مقاييس الموقع كالوسط الحسابي والوسيط إلا أن الظاهرتين مختلفتان. فمثلا، إذا كانت درجات الحرارة في بلد ما، خلال الليل والنهار، هي:

$28^0, 32^0, 23^0, 35^0, 24^0, 34^0$ فيكون معدل درجات الحرارة 29.3^0 وإذا كانت درجات الحرارة في بلد آخر هي: $17^0, 40^0, 19^0, 42^0, 20^0, 38^0$ فإن معدل درجات الحرارة فيها 29.3^0. وهذا يعني أن معدل درجات الحرارة في البلدين متساو.

ولو نظرت إلى مفردات البيانات في كل من البلدين لوجدت اختلافا بينهما، وهذا يعني أن الوسط الحسابي لا يكفي لوصف البيانات أو للحكم على تشابهها.

وكمثال آخر، صف دراسي فيه 50 طالبا معظم علاماتهم تقع ما بين 90 ,50 وعند حساب المعدل (الوسط الحسابي) وجد أنه يساوي 68، وصف آخر معظم العلامات فيه ما بين 67 ,70 ومعدله 68.

من الواضح وجود فروق بين هذين الصفين بالرغم من تساوي الوسطين الحسابين فيهما. إذن لا بد من استعمال مقاييس أخرى تبين مدى اختلاف البيانات فيما بينها ومدى التفاوت والتغير بين مفرداتها. فهل هي متقاربة من بعضها البعض أم متباعدة؟ ان مقاييس التشتت تجيب عن هذه التساؤلات.

ومن هذه المقاييس

(1) المدى والمدى الربيعي The Range and Quartile Range

تعريف (11)

يعرّف المدى في البيانات بأنه (الفرق بين أعلى قيمة وأصغر قيمة في البيانات) فإذا كان المدى صغيرا كانت البيانات محصورة في فترة قصيرة، وإذا كان المدى كبيرا كانت البيانات تقع ضمن فترة طويلة.

تعريف (12)

يعرّف المدى في البيانات المجمعة (التوزيع التكراري)، بأنه الفرق بين الحد الأعلى للفئة العليا والحد الأدنى للفئة الدنيا.

ومن هذا التعريف يظهر أن المدى لا يعتمد على جميع البيانات بل على أكبر قيمة وأصغر قيمة فقط. وهذا يقلل من أهميته، إذ قد يحدث أن تكون القيمتان المتطرفتان (أكبر قيمة وأصغر قيمة) قيمتين شاذتين عندئذ يكون المدى كبيرا بينما مفردات البيانات ليست متباعدة عن بعضها البعض، فمثلا إذا كانت علامات الصف الثاني الأساسي في مدرسة ما هي: ,70 ,69 ,30 ,74 ,65 ,78 ,67 ,72 ,64 73 ,100 ,70 ,65 ,67 ,71 ,65 ,69 ,65 ,71 فإن المدى هو 70 = 30 - 100 بينما معظم العلامات واقعة بين 64 ، 78 أي أنها متقاربة من بعضها البعض. وأنت تلاحظ من هذا المثال أن معظم علامات الصف كانت متقاربة إلى حدّ ما إلا أن المدى كان كبيرا. على ماذا تدل هذه الظاهرة؟ انها توضح أن المدى مقياس لا يعتمد عليه كثيرا لتحديد تشتت البيانات.

كيف يمكن التخلص من هذا النقص؟ إن إحدى الطرق للتخلص من ذلك هي حذف العلامات المتطرفة باعتبار أنها شاذة. احذف العلامتين 30 ,100 فيصبح المدى للبيانات الجديدة ؟ 14 = 64 - 78

إذا حذفت أعلى 25% من البيانات وأدنى 25% منها ثم حسبت المدى للبيانات الجديدة فإنك تحصل على المدى الربيعي.

تعريف (13)

المدى الربيعي هو الفرق بين الربيع الثالث والربيع الأول، أي $Q_3 - Q_1$

تعريف (14) :**نصف المدى الربيعي هو** $\frac{Q_3 - Q_1}{2}$

مثال (20)

إذا كانت أكبر قيمة في البيانات 100 وأقل قيمة 20 وكان الربيع الأول $Q_1 = 45$ والربيع الثالث $Q_3 = 87$ فما هو المدى وما هو المدى الربيعي.

الحل: المدى = أكبر قيمة - أصغر قيمة

أي 80 = 20 - 100

المدى الربيعي يساوي $Q_3 - Q_1$

أي 42 = 45 - 87

ونصف المدى الربيعي يساوي $\mathbf{21} = \frac{42}{2}$

مثال (21)

إذا كان الربيع الأول لمجموعة من العلامات 37 والربيع الثالث 92 وأعلى علامة 99 وأقل علامة 22. جد المدى والمدى الربيعي.

الحل: المدى = أعلى علامة - أصغر علامة

أي 77 = 22 - 99

المدى الربيعي هو $Q_3 - Q_1$

= 92 - 37 = 55

ونصف المدى الربيعي يساوي $\mathbf{27.5} = \frac{55}{2}$

(2) التباين والانحراف المعياري Variance and Standard Deviation

إن أحد مقاييس التشتت التي تخطر على البال هو مجموع انحرافات البيانات عن وسطها الحسابي، أي($\sum(x_i - \bar{x})$ ، لكن هذا المجموع يساوي صفراً دائما، ولذلك لابد

من حذف الإشارة السالبة لنحصل على مقياس ذي معنى.

وإحدى الطرق التي تزيل بها الإشارة السالبة هي بتربيع الانحرافات، ونحن نستعمل مربعات الانحرافات هذه في حساب التباين.

تعريف (13)

إذا كانت البيانات $X_1, X_2, \ldots, X_n$ تمثل عينة عشوائية،

يعرف التباين على أنه $S^2 = \frac{\sum_{I=1}^{n}(x_i - \bar{x})^2}{(n-1)}$

حيث $\bar{x}$ الوسط الحسابي للبيانات.

لاحظ أننا نقسم على (n - 1) على أساس أن البيانات التي لدينا هي مجرد عينة أخذت لتدرسها ومن ثم تعمم النتائج على المجتمع الذي أخذت منه وقد تمت القسمة على (n - 1) لكي يكون للتباين S^2 بعض الصفات الإحصائية المرغوب فيها.

مثال (22)

احسب التباين للبيانات 4, 8, 6, 7, 5, 8, 4, 9, 7, 2.

الحل: نجد الوسط الحسابي وهو

$$6 = \frac{4+8+6+7+5+8+4+9+7+2}{10}$$

نجد فروق البيانات عن الوسط الحسابي ثم نربع تلك الفروق ونجمعها ونقسم على (n - 1) فنجد التباين

$$S^2 = \frac{(4-6)^2+(8-6)^2+(6-6)^2+(7-6)^2+(5-6)^2+(8-6)^2+(4-6)^2+(9-6)^2+(7-6)^2+(2-6)^2}{9}$$

$$S^2 = \frac{4+4+0+1+1+4+4+9+1+1+16}{9}$$

$$= \frac{45}{9} = 5$$

لقد عرفنا التباين لمجموعة من البيانات، فما هو التباين في حالة التوزيعات التكرارية؟

تعريف (16)

إذا كانت مراكز فئات توزيع تكراري هي: $x_1, x_2, \ldots x_h$,

وكانت التكرارات المقابلة لها $f_1, f_2, \ldots f_h$,

فالتباين S^2 يكون

$$S^2 = \frac{\sum_{i=1}^{h}(x_i - \bar{x})^2 f_i}{n-1}$$

حيث $\bar{x}$ = الوسط الحسابي للتوزيع التكراري.

n = مجموع التكرارات.

مثال (23)

احسب التباين للتوزيع التكراري في الجدول التالي:

حدود الفئة	التكرار
30-34	6
35-39	5
40-44	12
45-49	9
50-54	8

الحل: نجد مراكز الفئات ونضرب كل مركز بالتكرار المقابل له لكي نحسب الوسط الحسابي ثم نجد انحرافات مراكز الفئات عن الوسط الحسابي. نربّع هذه الانحرافات ونضرب كلاً منها بالتكرار المقابل لها كما يظهر في الجدول التالي:

مركز الفئة x_i	التكرار f_i	$x_i f_i$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 f_i$

32	6	192	-11	121	726
37	5	185	-6	36	180
42	12	504	1	1	012
47	9	423	4	16	144
52	8	416	9	81	648
	40	1720			1710

نجد الوسط الحسابي $\bar{x} = \frac{1720}{40} = 43$

ونجد التباين $S^2 = \frac{1710}{39} = 43.85$

وكما هو واضح من تعريف التباين فإن وحدته هي مربع وحدة البيانات فإذا أردنا مقياساً للتشتت وحدته وحدة البيانات نأخذ الجذر التربيعي الموجب ونحصل على الانحراف المعياري كمقياس للتشتت.

تعريف (17)

الانحراف المعياري هو الجذر التربيعي الموجب للتباين أي

$$S = \sqrt{S^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

للبيانات الأولية،

أو
$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2 f_i}{n-1}}$$

للبيانات المجمعة (التوزيع التكراري ذي الفئات).

إن طريقة حساب التباين طويلة وتحتاج إلى حساب أرقام كثيرة، وبخاصة إذا احتوى الوسط الحسابي على كسور، لأن ذلك يستوجب أن تحتوي الانحرافات على كسور، وبالتالي فإننا نحتاج إلى تربيع تلك

الكسور. هناك طريقة مختصرة لحساب التباين صالحة للاستعمال بالآلة الحاسبة مباشرة وتعطى بالنظرية:

نظرية (1)

التباين للبيانات الأولية هو:

$$S^2 = \frac{1}{n-1}\left[\sum x_i^2 - n\bar{x}^2\right]$$

وللبيانات المجمعة في توزيع تكراري ذي فئات= $S^2 = \frac{1}{n-1}\left[\sum x_i^2 f_i - n\bar{x}^2\right]$

حيث $\bar{x}$ = الوسط الحسابي.

n = مجموع التكرارات.

مثال (42)

احسب التباين للتوزيع التكراري في مثال (23) بطريقة نظرية (1) ثم احسب الانحراف المعياري.

الحل: نرتّب الحل كما في الجدول التالي:

مركز الفئة x_i	التكرار f_i	$(x_i f_i)$	(x^2_i)	$x^2_i f_i$
32	6	192	1024	6144
37	5	185	1369	6845
42	12	504	1764	21168
47	9	423	2209	19881
52	8	416	2704	21632
المجموع	40	1720		75670

نلاحظ أن 43 = $\bar{x} = \frac{1720}{40}$

نعوّض في القانون فنجد التباين:

$$S^2 = \frac{1}{n-1}\left[\sum x_i^2 f_i - n\bar{x}^2\right]$$

$$= [75670 - 40 \times 43^2]\,\frac{1}{39}$$

$$= \frac{1710}{39} = 43.85$$

والانحراف المعياري هو

$S = \sqrt{S^2} = \sqrt{43.85} = 6.6$ تقريباً .

(3) الانحراف المتوسط Mean Deviation :

في دراستنا للتباين، كنا قد تخلصنا من الإشارة السالبة لبعض الانحرافات $(x_i - \bar{x})$ بوساطة التربيع، والآن يمكننا حذف الإشارة السالبة عن طريق أخذ القيمة المطلقة وفي هذه الحال يمكننا تعريف أحد مقاييس التشتت كما يلي :

تعريف (18) :

إذا كان لدينا البيانات $x_1, x_2, \ldots, x_n$ فإن انحرافها المتوسط يكون :

$$M.D. = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

حيث $\bar{x}$ الوسط الحسابي للبيانات.

مثال (25) :

أوجد الانحراف المتوسط للبيانات :

7, 6, 10, 8, 4, 5, 7, 9

الحل : $\bar{x} = \frac{7+6+10+8+4+5+7+9}{8}$

$$= \frac{56}{8} = 7$$

$$M.D. = \frac{|7-7|+|6-7|+|10-7|+|8-7|+|4-7|+|5-7|+|7-7|+|9-7|}{8}$$

$$= \frac{0+1+3+1+3+2+0+2}{8} = 1\frac{1}{2}$$

أمـا فـي حالـة البيانـات المجمعـة (التوزيـع التكـراري(فيكـون تعريـف الانحراف المتوسط كما يلي :

تعريف (19) :

إذا كانت مراكز فئات التوزيع التكراري $x_1, x_2, \dots, x_h$

وكانت التكرارات المقابلة لها $f_1, f_2, \dots, f_h$

فإن الانحراف المتوسط يكون :

$$M.D. = \frac{\sum_{i=1}^{h} |x_i - \bar{x}| f_i}{n}$$

حيث $\bar{x}$ = الوسط الحسابي ، n= مجموع التكرارات.

مثال (26) : احسب الانحراف المتوسط للتوزيع التكراري في الجدول التالي :

x_i	f_i
2	5
5	3
8	4

11	3

الحل : نرتب الحل كما في الجدول التالي :

x_i	f_i	$x_i f_i$	$x_i - \bar{x}$	$\lvert x_i - \bar{x} \rvert$	$\lvert x_i - \bar{x} \rvert . f_i$
2	5	10	-4	4	20
5	3	15	-1	1	03
8	4	32	2	2	08
11	3	33	5	5	15
	15	90			46

الوسط الحسابي : $\bar{x} = \frac{90}{15} = 6$

الانحراف المتوسط : $M.D. = \frac{46}{15} = 3.07$

وكما نلاحظ من التعريف فالانحراف المتوسط يعتمد على جميع مفردات البيانات وهو سهل التعريف سهل الحساب إلا أنه لا يخضع للعمليات الجبرية بسهولة حيث يجب تعديل الإشارة ويجب معرفة فئات البيانات إذا ما أردنا إيجاد قيمته، ويتضح هذا بعدم وجود طريقة جبرية لحساب الانحراف المتوسط للمجموعة الناتجة من دمج مجموعتين من البيانات إذا علم عدد مفردات كل منهما والوسط الحسابي والانحراف المتوسط لكل منهما : في هذه الحالة يجب معرفة جميع المفردات بعينها حتى نتمكن من حساب الانحراف المتوسط.

3 : 7 أثـر التحـويلات الخطيــة علــى كـل مــن مقــاييس النزعــة المركزية والتشتت.

(1) أثر التحويلات الخطية على مقاييس النزعة المركزية

التحويل الخطي هو كل اقتران من النوع:

$$y = f(x) = a x + b$$

حيث b, a ثابتان.

وتبرز الحاجة لدراسة أثر التحويلات الخطية على مقاييس النزعة المركزية والتشتت في حالة كون البيانات معطاة حسب نظام معين من المقاييس ثم حوّلت إلى مقياس آخر وطلب حساب مقاييس النزعة المركزية، والتشتت للبيانات الجديدة مباشرة بدلا من الرجوع إلى البيانات القديمة، وتحويلها، ثم حساب مقاييس النزعة المركزية والتشتت بعد التحويل. فمثلا، ربما تعطى درجات الحرارة بالمقياس المئوي وتكون مقاييس النزعة المركزية والتباين قد حسبت فإنك ستتمكن من حساب مقاييس النزعة المركزية والتشتت بالدرجات الفهرنهايتية باستعمال النظريات التي تعطيك أثر التحويلات الخطية على هذه المقاييس.

مثال (27)

إذا كان لدينا العلامات 7 ,6 ,8 ,10 ,4 ,0 ,3

فكم تصبح هذه العلامات تحت تأثير التحويل الخطي $y = 2x - 3$

الحل: العلامة 7 يقابلها العلامة $y = 2 \times 7 - 3 = 11$

سيعطيك الجدول التالي جميع قيم y المقابلة للعلامات.

3	0	4	10	8	6	7	x
3	-3	5	17	13	9	11	y

والآن مـا هـو الوسـط الحسـابي للعلامـات فـي المثـال السـابق ومـا هـو الوسط الحسابي للعلامات بعد التحويل؟

من تعريف الوسط الحسابي تجد:

$$\bar{x} = \frac{7+6+8+10+4+0+3}{7} = \frac{38}{7}$$

$$\bar{y} = \frac{11+9+13+17+5-3+3}{7} = 55/7$$

والآن لو عوضت قيمة $\bar{x}$ في التحويل المعطى لوجدت $\frac{55}{7}$=3 - $2 \times \frac{38}{7}$ =20

أي أن التحويل الخطي يؤثر على الوسط الحسابي بنفس الطريقة التـي يؤثر فيها على البيانات وتوضح لنا النظرية التالية أن التحويل الخطـي يؤثر على الوسيط والمنوال بنفس الطريقة التي يؤثر بها على البيانات أيضا.

نظرية (2)

إذا عدلت مجموعة من البيانات (الأولية أو التي في توزيع تكراري)

حسب التحويل الخطي (المعادلة). $y = ax + b$

حيث b, a ثابتان، فإن الوسط الحسابي والوسيط والمنوال تتأثر جميعها بنفس صيغة التحويل.

أي أنه إذا كان الوسط الحسابي $\bar{x}$ والوسيط M_x والمنوال D_x فإن المقاييس الجديدة تصبح:

$\bar{y} = a\bar{x} + b$ (1)

$M_y = aM_x + b$ (2)

$D_y = aD_x + b$ (3)

مثال (28)

إذا كان الوسط الحسابي لعلامات صف مكون من 60 طالبا هو 70، والوسيط 68 والمنوال 65 حوّلت جميع العلامات وفق المعادلة:

$y = 0.8 x + 20$

حيث x العلامة قبل التحويل، و y العلامة بعد التحويل.

اعتمادا على ذلك أجب عما يلي:

أ- ما قيمة الوسط الحسابي بعد التحويل؟

ب- ما قيمة الوسيط بعد التحويل؟

جـ- ما قيمة المنوال بعد التحويل؟

د- ما قيمة مجموع علامات الطلاب بعد التحويل؟

الحل:

أ- $\bar{y} = 0.8\ \bar{x} + 20$

إذن $\bar{y} = 0.8 \times 70 + 20 = 76$

ب- الوسيط بعد التحويل يساوي

$0.8 \times 68 + 20 = 74.4$

جـ- المنوال بعد التحويل يساوي

$0.8 \times 65 + 20 = 72$

د- مجموع العلامات بعد التحويل =

عدد الطلبة × الوسط الحسابي بعد التحويل = 76 × 60 = 4560

(2) أثر التحويلات الخطية على مقاييس التشتت

في دراستنا مقاييس التشتت كان واضحا أن هذه المقاييس تصف التغير والتباعد بين البيانات وأنها تعتمد على الانحرافات فيما بينها أو على الانحرافات عن الوسط الحسابي. بذلك فإننا نتوقع أن لا تتأثر هذه المقاييس بإضافة أي قيمة، موجبة كانت أم سالبة، إلى البيانات الأولية، وذلك لأن إضافة أي عدد إلى قيمة كل مفردة من البيانات الأولية لا يكون لها أثر بعد حساب الانحرافات.

مثال (29)

إذا كان لديك العلامات

20, 19, 28, 6, 18, 21, 7, 23

(أ) أضفت 2 لكل علامة، فهل يتغير المدى؟

الحل: المدى قبل الإضافة يساوي 22 = 6 - 28

العلامات بعد الإضافة تصبح 25 ,9 ,23 ,20 ,8 ,30 ,21 ,22

إذن المدى بعد الإضافة يساوي 22 = 8 - 30

أي أن المدى لا يتغير بإضافة عدد ثابت إلى جميع البيانات.

(ب) ضرب كل علامة بالعدد 3، ما هو المدى بعد التعديل

العلامات بعد التعديل (الضرب في 3) هي: 57 ,60 ,69 ,21 ,63 ,54 ,18 ,84

المدى بعد الضرب يساوي 66 = 18 - 84

أي أنه يساوي 3 × المدى قبل التعديل

نظرية (3)

إذا ضربت كل مفردة من البيانات في عدد ثابت a وجمع لها عددا ثابت b أي إذا أجري التحويل $y = ax + b$ فإن :

أ) تباين البيانات الجديدة يساوي تباين البيانات الأصلية مضروبا في a^2

ب) الانحراف المعياري للبيانات الجديدة يساوي الانحراف المعياري للبيانات الأصلية مضروبا في $|a|$.

جـ) الانحراف المتوسط للبيانات الجديدة يساوي الانحراف المتوسط للبيانات الأصلية مضروباً في $|a|$.

(لاحظ أن $|a|$ هي القيمة المطلقة للعدد a أي أنها قيمة a بدون إشارة.

د) المدى للبيانات الجديدة يساوي المدى للبيانات الأصلية مضروبا في $|a|$.

هـ) المدى الربيعي للبيانات الجديدة يساوي المدى الربيعي للبيانات الأصلية مضروبا في $|a|$.

لاحظ أنه لم يكن هناك أي أثر على التباين أو الانحراف المعياري أو المدى الربيعي من جراء جمعك العدد b لكل مفردة من البيانات.

مثال (30)

إذا كان تباين مجموعة من البيانات 25 وانحرافها المتوسط 7 ، وكان المدى يساوي 22 وقد تم تحويل هذه البيانات حسب المعادلة y = 2 x + 10 فما هو تباين البيانات الجديدة؟ وما هو انحرافها المعياري وما هو الانحراف المتوسط وما هو المدى لها؟

الحل: إن التباين والانحراف المعياري والانحراف المتوسط والمدى لا تتأثر من جراء جمع العدد 10 إلى البيانات الأصلية ولكنها تتأثر بالضرب بالعدد 2 حيث التحويل حسب y = 2 x + 10.

التباين للبيانات الجديدة = 2^2x تباين البيانات الأصلية أي 25 × 4 = 100

الانحراف المعياري للبيانات الجديدة هو $\sqrt{100}$ 10 =أو 10 = $2 \times \sqrt{25}$

الانحراف المتوسط للبيانات الجديدة هو 14 = 7 × | 2|

المدى للبيانات الجديدة |2|x المدى للبيانات الأصلية ويساوي 44 = 22 × 2

مثال (31)

إذا كان المدى الربيعي لمجموعة من البيانات 20 وكان تباين البيانات 16، وقمت بضرب كل مفردة من البيانات في العدد 3 فما هو المدى الربيعي والتباين والانحراف المعياري للبيانات الجديدة؟

الحل: المدى الربيعي للبيانات الجديدة = المدى الربيعي للبيانات الأصلية مضروبا في |3| أي 20 × 3 = 60

تباين البيانات الجديدة = 16 × 3^2 = 144

الانحراف المعياري للبيانات الجديدة = 12 = $\sqrt{144}$

مثال (32)

عدّلت علامات مجموعة من الطلبة حسب المعادلة y = 0.7x + 30

فإذا كان الانحراف المعياري للعلامات 8 فما هو الانحراف المعياري بعد التعديل؟

الحل: الانحراف المعياري بعد التعديل =

الانحراف المعياري للبيانات الأصلية مضروبا في |0.7| ويساوي

8 × |0.7| = 5.6

3 : 8 وصف البيانات بطريقة الصندوق والطرفين Box - and whiskers plot

بعد دراستنا للتوزيعات التكرارية وعرضها بيانياً بطريقة المدرج التكراري والمضلع التكراري تعرفنا على أشكال هذه التوزيعات وصفاتها من حيث التماثل والالتواء ووجود المنوالات.

والآن، بعد دراستنا لمقاييس النزعة المركزية ومقاييس التشتت نعطي طريقة بيانية لوصف التوزيعات وهي طريقة الصندوق والطرفين وتتلخص بما يلي:

1- نجد قيم خمسة مقاييس للتوزيع وهي: القيمة الصغرى، القيمة العظمى، الربيع الأول Q_1، الربيع الثاني (الوسيط) Q_2، والربيع الثالث Q_3. وتعتبر هذه القيم الخمس من أهم المقاييس لوصف التوزيع وتسمى "ملخص الخمس النقاط" للتوزيع.

2- نرصد النقاط الخمسة على خط أفقي (أو عمودي) ونرسم مستطيلاً قاعدتهُ الفترة (Q_1, Q_3) وعرضه بطول مناسب. ومن نقطة الوسيط نرسم مستقيما موازياً لضلع العرض فيصبح المستطيل منقسماً إلى مستطيلين متلاصقين.

نرسم الطرفين وهما الخط الموازي لقاعدة المستطيل والواصل من منتصف عرض المستطيل إلى أصغر قيمة في البيانات والطرف الثاني هو الخط الواصل إلى القيمة العظمى في البيانات كما هو موضح في المثال التالي:

مثال (33)

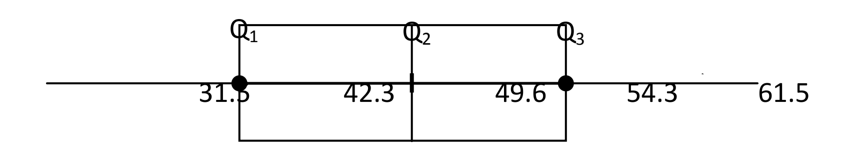
أوجد ملخص النقاط الخمس للتوزيع التكراري للزمن الذي استغرقه 50 طالباً للإجابة عن أسئلة امتحان شهري المعطى في مثال (17).

ثم أرسم الصندوق والطرفين لذلك التوزيع.

الحل: كما جاء في حل مثال (17) فإن:

$Q_1 = 42.3$ ،$Q_2 = 49.6$ ،$Q_3 = 54.3$

ومن جدول التوزيع التكراري نلاحظ أن القيمة الصغرى 31.5 (الحد الأدنى الفعلي للفئة الدنيا) والقيمة العظمى 61.5 (الحد الأعلى الفعلي للفئة العليا). والآن نرسم الصندوق والطرفين كما في الشكل (2) :



الشكل (2)

إن العرض البياني بطريقة الصندوق والطرفين يوضّح لنا فيما إذا كان التوزيع شبه متماثل أو أنه ملتو نحو اليمين أو ملتو نحو اليسار، فإذا ظهر أن المستطيل على Q_1Q_2 مساو للمستطيل على Q_2Q_3، وأن الطرف الواصل إلى الحد الأعلى مساو للطرف الواصل إلى الحد الأدنى. يكون التوزيع متماثلاً. أما إذا كان المستطيل على Q_1Q_2أكبر من المستطيل على Q_2Q_3وكان الطرف الأيسر أطول من الطرف الأيمن فيتضح أن التوزيع ملتو نحو اليسار. وبنفس الطريقة نحدد فيما إذا كان التوزيع ملتوياً نحو اليمين.

إضافة إلى فائدة رسم الصندوق والطرفين بإعطاء فكرة واضحة عن شكل التوزيع التكراري من حيث التماثل والالتواء فإن فائدة هذا الرسم تكمن في المقارنة بين توزيعين تكراريين (مجموعتين من البيانات) أو أكثر وذلك برسم الصندوق والطرفين لكل منهما والمقارنة بينهما عن طريق هذين الرسمين.

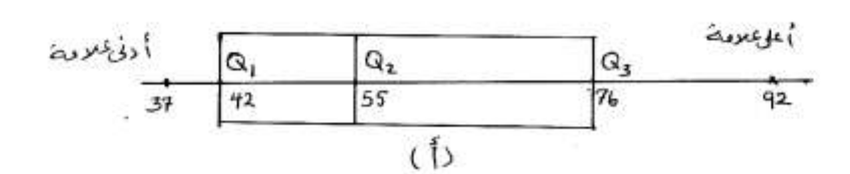
لاحظ أنه يمكن رسم الصندوق والطرفين على خط عمودي بدلاً من خط أفقي.

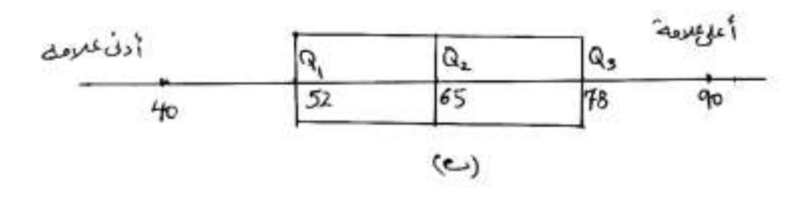
مثال (34) :

أعطت نتائج امتحان مادة اللغة الفرنسية للصفين أ, ب في إحدى المدارس ما يلي، قارن بين نتائج الصفين.

لصف	أدنى علامة	أعلى علامة	Q_1	Q_2	Q_3
أ	37	92	42	55	76
ب	40	90	52	65	78

الحل: نرسم الصندوق والطرفين لكل من الصفين أ، ب كما في الشكل (3).





الشكل (3)

نلاحظ أن نتائج الصف أ ملتوية نحو اليمين حيث يظهر في الرسم أن طول المستطيل على Q_2Q_3 أكبر من الطول Q_1Q_2 ونلاحظ أن نتائج الصف ب تكوّن توزيعاً متماثلا تقريباً.

3 : 9 معامل التغيّر Coefficient of Variation

إن مقاييس التشتت التي درسناها تعتمد كلها على الوحدة المستعملة في البيانات ولكي نحصل على مقياس لا يعتمد على الوحدة المستعملة نعرّف معامل التغيّر.

تعريف (20) :

للبيانات في عيّنة ما نعرف معامل التغير (بأنه $C.V. = \frac{S}{\overline{X}} \times 100\%$

إن أهم استعمالات معامل التغير هو للمقارنة بين التغير في عدة مجموعات من البيانات أو عدة توزيعات تكرارية ولا فرق في ذلك إذا كانت الوحدات المستعملة مختلفة أو هي ذاتها.

مثال (35):

أعطت العينة (أ) لعلامات مجموعة من الطلبة ما يلي:

الوسط الحسابي $\overline{X}_1 = 70$ ، الانحراف المعياري $S_1 = 12$

وأعطت العينة (ب) للمبيعات اليومية بالدينار لعدد من المحلات التجارية

$\overline{X}_2 = 6500$, $S_2 = 500$

أي العينتين أكثر تغيراً.

الحل: مع أن الوحدات المستعملة في العينتين مختلفة عن بعضها، فهي في (أ) علامة وفي (ب) دينار إلا أننا نتمكن من معرفة أيها أكثر تغيراً بحساب معامل الاختلاف لكل عينة.

عينة (أ)
$$C.V. = \frac{12}{70} \times 100\% = 17.14\%$$

العينة (ب)
$$C.V. = \frac{500}{6500} \times 100\% = 7.69\%$$

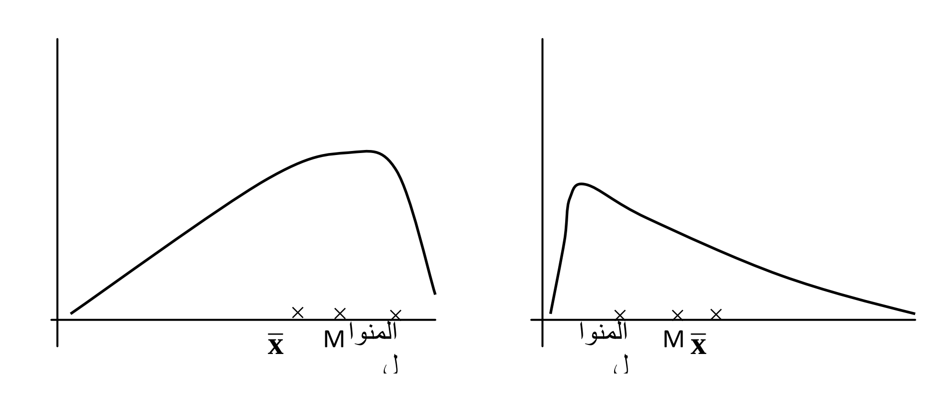
إذن العينة (أ) أكثر تغيراً.

3 : 10 مقاييس الالتواء Coefficients of Skewness

لقد مر معنا في وصف أشكال التوزيعات التكرارية كيف نتعرف على التواء التوزيع فيكون التوزيع ملتويا نحو اليمين أو موجب الالتواء إذا كان ممتدا أكثر نحو اليمين أما إذا كان له طرف واحد ممتد نحو اليسار فيقال إنه سالب الالتواء.

عندما يكون التوزيع ملتويا نحو اليمين فإن القيم المتطرفة نحو اليمين تؤثر على الوسط الحسابي وتسحبه نحو اليمين وبذلك يكون الوسط الحسابي أكبر من الوسيط. أما إذا كان التوزيع ملتويا نحو اليسار فإن الوسط الحسابي يكون أصغر من الوسيط.

والأشكال التالية تمثل نماذج من هذه التوزيعات :



ويمكن استعمال التعريف التالي لنقيس الالتواء ونعرف إشارته.

تعريف (21) :
يعرف مقياس الالتواء لمجموعة من البيانات أو أكبر قليلاً لتوزيع تكراري بالمعادلة.

$$\gamma_1 = \frac{3(\bar{x} - M)}{S}$$

$\bar{x}$: الوسط الحسابي.

M : الوسيط.

S : الانحراف المعياري.

ويستفاد من مقياس الالتواء في معرفة نوعية الالتواء للتوزيع التكراري، فإذا كان الالتواء موجبا فهذا يعني أن الوسط الحسابي أكبر من الوسيط وأن للتوزيع طرفا ممتدا إلى اليمين أكثر من الطرف الثاني الممتد إلى اليسار. أما إذا كان مقياس الالتواء سالبا فهذا يعني أن الالتواء نحو اليسار والطرف الأيسر هو الأكثر امتدادا.

وهناك فائدة ثانية وهي للمقارنة بين توزيعين تكراريين أو مجموعتين من البيانات. فالمجموعة التي تكون القيمة المطلقة لمقياس الالتواء لها أكبر يكون توزيعها ملتويا أكثر من توزيع المجموعة الأخرى مع الأخذ بعين الاعتبار أن إشارة مقياس الالتواء تعين اتجاه الالتواء. فمثلا إذا كان مقياس الالتواء لمجموعة من البيانات (أ) هو 0.57 ولمجموعة أخرى (ب) هو 0.63- فهذا يعني أن التواء المجموعة (ب) أكبر من التواء المجموعة (أ) لأن $|0.57| < |-0.63|$ بينما التواء (أ) لليمين والتواء (ب) إلى اليسار.

مثال (36) :
توزيع تكراري فيه الوسط الحسابي $\bar{x} = 50$ والوسيط M = 48 والتباين $S^2 = 144$ ما مقياس الالتواء فيه :

الحل :

$$\gamma = \frac{3(\bar{x} - M)}{S}$$
$$= \frac{3(50 - 48)}{12} = 0.5$$

أي أنه ملتو نحو اليمين والتواؤه يساوي 0.5.

أما القسمة على الإنحراف المعياري في تعريف مقياس الالتواء فذلك لجعله غير معتمد على وحدة القياس المستعملة في البيانات.

ويستعمل في العادة مقياس آخر للالتواء يعتمد تعريفه على العزم الثالث حول الوسط الحسابي ولذلك نعرف العزوم أولا.

تعريف (21) :

العزم k للبيانات $x_1, x_2, \ldots, x_n$ حول وسطها الحسابي هو :

$$m_k = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^k}{n}$$

والعزم k للتوزيع التكراري ذي الفئات حول وسطه الحسابي هو :

$$m_k = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^k f_i}{n}$$

تعريف (22) :

مقياس الالتواء لتوزيع تكراري أو مجموعة من البيانات هو النسبة بين العزم الثالث حول الوسط الحسابي ومكعب الانحراف المعياري، أي :

$$\gamma_2 = \frac{m_3}{S^3}$$

وهذا المقياس لا يعتمد على وحدة قياس البيانات.

هناك مقاييس أخرى للالتواء منها :

معامل الالتواء الربيعي (Quartile Coefficient of Skewness) :

ويستعمل الربيعات في تعريفه وهو :

$$\gamma_3 = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1} = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}$$

ومنها أيضا معامل الالتواء المئيني (10-90 Percentile Coefficient of Skewness):

ويستعمل المئينات في تعريفه وهو :

$$\gamma_4 = \frac{(P_{90} - P_{50}) - (P_{50} - P_{10})}{P_{90} - P_{10}} = \frac{P_{90} - 2P_{50} + P_{10}}{P_{90} - P_{10}}$$

مثال (37) :

احسب مقياس الالتواء للتوزيع التكراري في مثال (23).

الحل : بالرجوع إلى جدول الحل لمثال (12) فنعيد كتابة الأعمدة الأربعة الأولى لنكمل عليها ونجد العزم الثالث حول الوسط الحسابي فنجد :

x_i مركز الفئة	f_i التكرار	$x_i f_i$	$(x_i - \bar{x})$	$(x_i - \bar{x})^3$	$(x_i - \bar{x})^3 . f_i$
32	6	192	-11	1331	7986
37	5	185	-6	216	1080
42	12	504	1	1	12
47	9	423	4	64	576
52	8	416	9	729	5832
					15486

العزم الثالث حول الوسط الحسابي : $m_3 = \frac{15486}{40} = 387.15$

معامل الالتواء : $\gamma_2 = \frac{m_3}{S^3} = \frac{387.15}{(\sqrt{43.85})^3} = \frac{387.15}{290.37} = 1.33$

مثال (38) :احسب معامل الالتواء الربيعي لعلامات الصف (أ) والصف (ب) في مثال (34) :

$$\gamma_3 = \frac{(76-55)-(55-42)}{(76-42)} = \frac{8}{34} = 0.23$$

معامل الالتواء الربيعي للصف (ب) يساوي :

$$\gamma_3 = \frac{(78-65)-(65-52)}{(78-52)} = \frac{0}{26} = 0$$

ومنـه يظهـر أن توزيـع علامـات الصـف (أ) ملتـو نحـو اليمـين بينمـا توزيع علامات الصف (ب) متماثل ، كما جاء في مثال (34).

تمارين

1-3 : ذكرت نشرة الأحوال الجوية أن معدلات درجات الحرارة في مدينـة عمـان خلال أسبوعين بالدرجات المئوية (سيلشيوس) كما يلي :

27, 23, 20, 17, 12, 22, 13, 10, 9, 14, 18, 25, 26, 23

أ) أوجد الوسط الحسابي (المعدل) لدرجات الحرارة.

ب) أوجد الوسيط.

2-3 : يمثل الجدول التالي الودائع التي وضعت في أحد البنوك في مدة شهر.

حدود الفئات بالدنانير	**التكرار**
200-499	15
500-700	12
800-1099	18
1100-1399	15
1400-1999	20
2000-3999	10

أ- أوجد الوسط الحسابي للودائع.

ب- أوجد الوسيط.

جـ- أوجد المئين الخامس والعشرين (أي Q_1).

د- أوجد المئين الخامس والسبعين (أي Q_3).

هـ- أوجد معامل الالتواء الربيعي.

3-3 : يمثل الجدول التالي التوزيع التكراري لأوزان أمتعة مجموعة من المسافرين جوا.

حدود الفئات بالكغم	التكرار
10-15	20
16-21	18
22-27	16
28-33	14
34-39	12
40-45	20

أ- أوجد الوسط الحسابي لأوزان الأمتعة.

ب- أوجد الوسيط.

جـ- أوجد الربيع الأول Q_1 والربيع الثالث Q_3.

د- أوجد المئين التسعين والمئين العاشر.

هـ- أوجد معامل الالتواء الربيعي ومعامل الالتواء المئيني.

4-3 : يمثل الجدول التالي التوزيع التكراري لأعمار بطاريات السيارات (لأقرب شهر) التابعة لإحدى المؤسسات.

المركز	التكرار
22	17
27	12
32	10
37	11
42	04

أ- أوجد الوسط الحسابي لأعمار البطاريات.

ب- أوجد المئين الثمانين.

جـ- أوجد الوسيط.

5-3 : تمثل العلامات في تمرين (2-5) علامات 36 طالبا في أحد المقررات :
أوجد الوسط الحسابي والوسيط.

6-3 : في تمرين (3-1) ، أوجد التباين والانحراف المعياري والانحراف المتوسط للبيانات، ومقياس الالتواء γ_2.

7-3 : في تمرين (3-2) (أ) أوجد الانحراف المعياري للودائع.

(ب) أوجد معامل التغير للودائع.

(جـ) اكتب ملخص الإحصاءات الخمس للودائع ثم أرسم الصندوق والطرفين.

8-3 : أجب عن الأسئلة السابقة (أ)، (ب)، (جـ) لتوزيع أوزان الأمتعة المعطى في تمرين (3-3).

9-3 : أوجد التباين والانحراف المعياري والانحراف المتوسط لتوزيع أعمار البطاريات في تمرين (3-4).

10-3 : أوجد التباين واكتب ملخص الإحصاءات الخمس لعلامات الطلبة في تمرين (3-5)، ثم أرسم الصندوق والطرفين وأدرس شكل التوزيع من ذلك الرسم.

11-3 : أعطت العينتان (أ)، (ب) المقادير التالية:

(أ) $$\sum_{i=1}^{40} x_i = 240, \quad \sum_{i=1}^{40} x_i^2 = 2050$$

(ب) $$\sum_{i=1}^{30} y_i = 210, \quad \sum_{i=1}^{30} y_i^2 = 1740$$

أوجد الوسط الحسابي والانحراف المعياري لكل عيّنة، ثم أوجد أي العينتين أكثر تغيراً.

12-3 : الجدول التالي يعطي ملخص إحصاءات عن الرواتب الشهرية بالدينار لأعضاء هيئة التدريس في إحدى الجامعات:

عدد أعضاء هيئة التدريس	الوسط الحسابي	الوسيط	الانحراف المعياري
110	815	850	100

أ. ما مجموع الرواتب التي تدفعها الجامعة لأعضاء هيئة التدريس في السنة؟

ب. إذا أعطت الجامعة علاوة إضافية للنقل مقدارها 40 ديناراً لكل عضو هيئة تدريس فما هو الوسط الحسابي والوسيط والانحراف المعياري للرواتب الجديدة مع علاوة النقل الإضافية؟

جـ. أوجد معامل التغير قبل علاوة النقل وبعدها.

13-3 : يمثل الجدول التالي التوزيع التكراري للرصيد في الحساب الجاري لعدد من الطلبة لشهر شباط 1999.

التكرار	حدود الفئات (لأقرب دينار)
300	60-69
450	70-79
680	80-89
720	90-99
150	100-109

أ. أوجد الانحراف المتوسط للأرصدة.

ب. أوجد الانحراف المعياري للأرصدة.

جـ. أوجد مقياس الالتواء γ_2.

الفصل الرابع
الاحتمال Probability

4 : 1 مقدمة

قد تكون سمعت أحد زملائك يقول "احتمال نجاحي في مقرر الإحصاء 0.9" وآخر يقول إن احتمال وصول معظم الدارسين في الوقت المحدد ضئيل ولا يزيد عن 0.1 لأن الأمطار غزيرة. لو تمعنت في معنى هذه العبارات ومثيلاتها لاستطعت أن تستنتج أن احتمال حادث ما هو عدد حقيقي لا يزيد على 1 ولا ينقص عن الصفر. ولكن كيف نعيّن هذا العدد؟

هناك طرق متعّددة لتعيين هذا العدد، الاحتمال، منها :

أولاً: طريقة الرأي الشخصي Personal Probalility

وهي تعتمد على التقدير الشخصي للفرد الذي يرغب في تعيين إمكانية حدوث حادث أي احتمال ذلك الحادث كما ذكرنا آنفاً. فالشخص الذي يجد نفسه قد فهم المادة في مقرر ما وأنه بذل جهداً في دراستها قد يعيّن احتمال 0.9 مثلاً لحصوله على علامة امتياز في ذلك المقرر.

ثانياً: تعريف الاحتمال بالتكرار النسبيRelative Frequency

من طرق تعريف الاحتمال تعريفه بالتكرار النسبي، وهو تعريف شائع الاستعمال، أساسه المشاهدات والبيانات التي يجمعها المرء من المحاولات المتكررة (بشكل لا نهائي) للتجربة الإحصائية تحت الدراسة، أي أنه مبني على فكرة التكرار النسبي.

تعريف (1)

يتلخص تعريف الاحتمال، بمفهوم التكرار النسبي، في أنه: إذا تكرر إجراء تجربة إحصائية n من المرات تحت نفس الظروف وكان عدد المرات التي تؤدي إلى الحادث A يساوي n(A)، فإن التكرار النسبي لهذا الحادث هو $\frac{n(A)}{n}$.

إذا كبرت n بدون حدود، وكانت n(A) تكبر معها بحيث يؤول التكرار النسبي في النهاية إلى عدد ثابت، وليكن P فإنك تقول إن احتمال الحادث A هو P.

وتسهيلا عليك في أن تستوعب معنى هذا التعريف، قم برمي قطعة نقود عددا معينا من المرات وسجل عدد المرات التي تظهر فيها الصورة H. احسب التكرار النسبي لظهور الصورة. كرر العملية عددا أكثر من المرات واحسب التكرار النسبي لظهور الصورة, ثم أعد العملية مرة ثانية وسجل التكرار النسبي. ماذا تتوقع؟ إذا كانت قطعة النقود منتظمة فإنك تتوقع أن يقرب التكرار النسبي من $\frac{1}{2}$ عندما تجري هذه التجربة (رمي قطعة النقود) عددا كبيرا جدا من المرات.

أي أنه عندما يكبر عدد الرميات بدون حدود فإن عدد ظهور الصورة يكبر والنسبة تؤول إلى عدد ثابت وهو $\frac{1}{2}$. ونقول إن احتمال ظهور الصورة (H) عند رمي قطعة النقود المنتظمة = $\frac{1}{2}$.

لعلّك تلاحظ عند قيامك بإجراء مثل هذه التجربة مراراً وتكراراً فإنك عمليا لا تستطيع إجراء التجربة عددا لا نهائيا من المرات ولذلك نكتفي بتعريف الاحتمال بمفهوم التكرار النسبي على أنه: عدد المرات التي حدث فيها الحادث مقسوما على جميع المرات التي أجريت بها التجربة.

مثال (1):

لتحديد احتمال أن تكون كمية الأمطار أكثر من قيمة معينة، قام باحث بدراسة معدل نزول المطر طيلة العشرين سنة الماضية فوجد أنه كان هناك 7 سنوات زادت كميات الأمطار في كل منها عن 400ملم ووجد أن كمية الأمطار أعلى من 450ملم في ثلاث سنوات.

(أ) ما احتمال أن تكون كميات الأمطار للعام القادم أكثر من 400 ملم؟

(ب) أكثر من 450 ملم.

الحل: بالعودة إلى التعريف (1) فإن احتمال أن تزيد كمية الأمطار عن 400 ملم هو: $\frac{7}{20}$ 0.35 =

واحتمال أن تزيد الكمية عن 450 ملم هو 0.15 = $\frac{3}{20}$ هناك بعض

الاعتراضات على تعريف الاحتمال بالتكرار النسبي منها أنك لا تستطيع عمليا إعادة التجربة عددا غير محدود من المرات، كما أنك لا تعلم أن ستكون هناك نهاية واحدة للتكرار النسبي أم لا. أي أنك لا تعلم إذا كان التكرار النسبي سيقرب من عدد ثابت أم لا.

هذه ناحية، ومن ناحية أخرى، فهناك إيجابيات لهذا التعريف، منها اعتماده على خاصية الثبات الإحصائي، التي ربما تظهر بعد إجراء التجربة عددا محدودا من المرات فيتضح للباحث أن النتيجة تثبت عند قيمة معينة، وهذه خاصية مفيدة إذا لم تكن تعلم شيئاً عن احتمال ظاهرة معينة. عند ذلك الجأ إلى استعمال بيانات إحصائية كافية تستطيع بناءً عليها، وباستعمال التكرار النسبي تعريف احتمال

حدوث تلك الظاهرة. فمثلاإذا كانت نسبة مالكي السيارات الخاصة في بلد معين تساوي 0.4 فهذا يعني أن التكرار النسبي لمالكي السيارات الخاصة يساوي 0.4. وبالتالي، إذا أردت معرفة احتمال أن يكون أحد القاطنين في تلك البلد مالكا لسيارة خاصة فإن بإمكانك القول أن ذلك الاحتمال يساوي 0.4.

مثال (2) :

إذا كان طلبة إحدى الكليات موزعين على التخصصات المختلفة فيها حسب الجدول (1) وقمت أنت بمقابلة أحد الطلبة، فما هو احتمال أن يكون الطالب في تخصص المحاسبة؟

جدول (1)

التخصص	عدد الطلبة
إدارة أعمال	320
محاسبة	480
اقتصاد	300
علوم مالية ومصرفية	500

الحل: التكرار النسبي لطلبة تخصص المحاسبة يساوي

$$\frac{480}{320+480+300+500} = 0.3$$

إذن احتمال أن يكون الطالب في تخصص المحاسبة يساوي 0.3

وبسبب الاعتراضات على تعريف الاحتمال بالتكرار النسبي وتسهيلا علينا في استعمال الاحتمالات بشكل مفيد في الفصول القادمة فإننا سنعرف الاحتمال بشكل رياضي، علما بأن تعريف الاحتمال بالتكرار النسبي يحقق الشروط التي سنقدمها في التعريف الرياضي والذي يعتمد على دراسة الفضاء العيني والحوادث والتي بدورها تعتمد على دراسة المجموعات.

2 : 4 المجموعات (Sets)

إن دراسة المجموعة ذات أهمية كبيرة وضرورية لدراسة الاحتمال، والمجموعة هي تجمع أي عدد من العناصر أو الأشياء وسنمثل المجموعات بحروف لاتينية كبيرة مثل ... ,A, B, C وعناصر المجموعة بحروف صغيرة مثل ... ,a, b, c. يستعمل الرمز ∈ ليعني "ينتمي إلى" والرمز ∉ ليعني "لا ينتمي إلى" وهكذا فإن $a \in A$ تعني أن العنصر a ينتمي إلى المجموعة A ، أما $b \notin A$ فيعني أن العنصر b لا ينتمي إلى المجموعة A.

وعند دراسة أي مجموعة يجب التأكد من أنها معرفة تماما وذلك يعني أنه إذا أعطينا أي عنصر فإنه سيكون بإمكاننا معرفة فيما إذا كان هذا العنصر منتميا للمجموعة أو غير منتم إليها

هناك طريقتان لوصف المجموعات :

1- طريقة العد (Roster Method)

وذلك بوضع جميع عناصر المجموعة بين حاصرتين.

مثل $A = \{1, b, c, d, e, f\}$

أو $B = \{H, T\}$

أو $S = \{1, 2, 3, ... 30\}$

2- طريقة القانون (Rule Method) :

وذلك بوصف المجموعة بقانون يوضع بين حاصرتين، فمثلا :

A = {x | x عدد صحيح موجب أقل من 9}

أو B = {x| x هو قسم علمي في جامعة اليرموك}.

تعريف (2) : المجموعة الخالية (Empty set)

هي المجموعة التي لا يوجد فيها عناصر مطلقا ونعبر عنها بالرمز ϕ أو { }.

تعريف (3) : المجموعة الجزئية (Subset)

نقول إن A مجموعة جزئية من B ونرمز لها $A \subset B$ إذا كان كل عنصر في A منتميا للمجموعة B.

تعريف (4) : التساوي (Equality) :

نقول إن المجموعة A تساوي المجموعةB أي $A = B$ إذا كان B , A تحتويان على نفس العناصر، ويكون ذلك إذا تحقق الشرطان $A \subset B$ و $B \subset A$.

تعريف (5) : المجموعة الكلية (Universal Set) :

هي المجموعة التي تدرس جميع المجموعات تحت البحث باعتبارها مجموعات جزئية منها .

العمليات الجبرية على المجموعات (Set Operations)

1- الاتحاد (Union) :

اتحاد المجموعتين A و B أي $A \cup B$ هو مجموعة العناصر الموجودة في A أو في B أو في كليهما. وبعبارة أخرى.

$$A \cup B = \{x | x \in B \text{ أو } x \in A\}$$

2- التقاطع (Intersection)

تقاطع المجموعتين B و A أي $A \cap B$ هو مجموعة العناصر المشتركة بين A و B أي هو مجموعة العناصر المنتمية إلى A وإلى B في نفس الوقت.

$$A \cap B = \{x | x \in B \text{ و } x \in A\}$$

وسنستعمل أحيانا الرمز AB للدلالة على $A \cap B$.

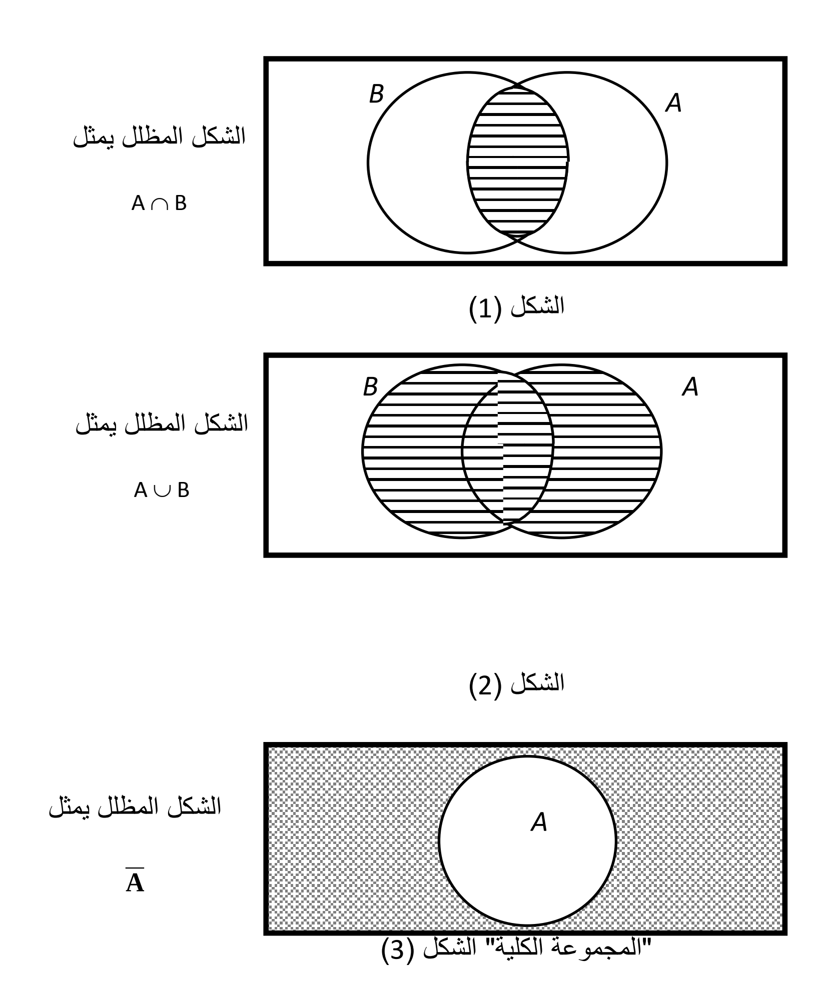
3- المتممة (Complement) :

متممة A (بالنسبة إلى المجموعة الكلية S) هي مجموعة العناصر المنتمية إلى المجموعة الكلية S ولا تنتمي إلى A , ونمثلها بالرمز $\overline{A}$ وبعبارة أخرى :

$$\overline{A} = \{x | x \in S \,;\, x \notin A\}$$

تعريف (6) : إذا لم يوجد عناصر مشتركة بين مجموعتين نقول إن المجموعتين منفصلتان عن بعضهما البعض أي أنه إذا كان $A \cap B = \phi$ فإن A , B منفصلتان (disjoint).

يمكننا تمثيل المجموعات والعمليات عليها باستعمال أشكال هندسية وذلك بالتعبير عن المجموعة الكلية بمستطيل أو مربع ثم تمثيل أي مجموعة جزئية للمجموعة الكلية بشكل هندسي معين (كالدائرة أو المثلث أو غيرهما) يرسم داخل المستطيل، وتكون هذه الأشكال توضيحية توضح العمليات التي نجريها على المجموعات وتعرف هذه الأشكال بأشكال فان (Venn diagram) ومن الأمثلة على ذلك الأشكال الآتية :



الشكل (1)

الشكل (2)

الشكل (3)

نذكر باختصار فيما يلي خواص العمليات الجبرية على المجموعات :

I) خاصية التبديل Commutative Law

خاصية التبديل للاتحاد $A \cup B = B \cup A$

خاصية التبديل للتقاطع $A \cap B = B \cap A$

II) خاصية التجميع (Associative Law)

خاصية التجميع للاتحاد $A \cup (B \cup C) = (A \cup B) \cup C$

خاصية التجميع للتقاطع $A \cap (B \cap C) = (A \cap B) \cap C$

III) خاصية التوزيع (Distributive Law)

خاصية التوزيع للتقاطع على الاتحاد $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

خاصية التوزيع للاتحاد على التقاطع $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

IV) قوانين ديمورغان (De Morgan's Laws)

$$\left(\overline{A \cup B}\right) = \overline{A} \cap \overline{B}$$
$$\left(\overline{A \cap B}\right) = \overline{A} \cup \overline{B}$$

هذه القوانين تعطينا العلاقة بين عمليات الاتحاد والتقاطع وأخذ المتممة. ويمكن تعميم القوانين السابقة إلى n من المجموعات بدلا من مجموعتين.

3 : 4 التجربة الإحصائية والفضاء العيني والحوادث.

للإحصاء أهمية كبيرة في استقراء النتائج وصياغة التعميمات عن المجتمع بعد دراسة عينة عشوائية تؤخذ من ذلك المجتمع، وتستعمل نظرية الاحتمال في إعطاء التبريرات الرياضية للتوصل إلى هذه التعميمات من

دراسة العينات كما وأن هذه النظرية تساعد في تحديد مدى صدق تمثيل العينات للمجتمع ومدى الثقة في الاستدلال.

لو أمعنت النظر في البيانات التي تصادفك في الحياة لوجدتها مجموعة من مشاهدات وقياسات يسجلها العلماء والباحثون، نتيجة إجراء التجارب. لاحظ كاتب سجل المواليد في مستشفى معين تجد أنه يقوم بتسجيل نتائج تجربة.

ولو قذفت قطعة نقود إلى أعلى وشاهدت الوجه الذي يظهر بعد سقوطها وسجلت هذه المشاهدة فإنك تكون قد قمت بتسجيل نتائج تجربة.

وإذا وقفت عند إشارة ضوئية على تقاطع طرق وسجلت عدد السيارات المارة في اتجاه معين في كل دقيقة فإنك تكون قد قمت بتسجيل نتائج تجربة. إن التجارب السابقة ومثيلاتها تدعى تجارب إحصائية.

تعريف (7):
التجربة الإحصائية: هي أي عملية أو مجموعة عمليات محددة لا تعرف نتائجها مسبقا بشكل حتمي، أي لا يستطاع التنبؤ بنتائجها بشكل مؤكد. وبعبارة أخرى هي كل عملية تعطي مشاهدة أو قياسا لظاهرة.

فمثلا: إذا رميت زهرة نرد (زهر الطاولة) وسجلت العدد الذي يظهر فيه إلى أعلى، فهذه تجربة إحصائية، وإذا أخضعت مجموعة من الحيوانات كالأرانب أو الفئران لبعض الأدوية لتعرف تأثير تلك الأدوية عليها فتلك تجربة إحصائية أيضا.

ومن أهم الصفات التي يجب أن تتوفر في التجربة الإحصائية تحديد المشاهدات والبيانات التي تريد تسجيلها فيها.

ولكل تجربة إحصائية نتائج. وتعرف النتيجة للتجربة على أنها النتيجة البسيطة، أي النتيجة التي لا يمكن تقسيمها أو تحليلها إلى نتيجتين أو أكثر. وتسمى النتائج البسيطة التي يمكن الحصول عليها عند القيام بتجربة إحصائية، النتائج الممكنة الحدوث، وهي عناصر هامة في دراسة الاحتمالات.

مثال (3) :
إرم زهرة نرد مرّة واحدة، واكتب جميع النتائج البسيطة التي يمكن أن تحصل عليها.

الحل: من الواضح أن هناك ستة وجوه لزهرة النرد وهذه الوجوه تحمل الأعداد 1 ,2 ,3 ,4 ,5 ,6 إذن فمجموعة النتائج البسيطة الممكنة هي:

{6, 5, 4, 3, 2, 1}

تعريف (8) :
الفضاء العيني أو فضاء العينة Sample Space لتجربة إحصائية هو مجموعة جميع النتائج الممكنة لتلك التجربة.

ويعبر عن الفضاء العيني بالرمز Ω

مثال (4) :
اقذف قطعة نقود وارم زهرة نرد، مرة واحدة. ثم اكتب عناصر الفضاء العيني.

الحل: أرمز لوجه قطعة النقود الذي تظهر عليه الصورة بالحرف H، وللوجه الآخر، وهو الكتابة، بالحرف T.

من الواضح أن إحدى النتائج البسيطة هي ظهور الصورة على قطعة النقود والعدد 1 على زهرة النرد وبذلك يكون (H, 1) نتيجة بسيطة أي عنصرا في الفضاء العيني ولو حاولت أن تكتب جميع النتائج البسيطة التي يمكن الحصول عليها لوجدت أنها تمثل عناصر الفضاء العيني. وهي:

Ω = {(H, 1), (H, 2), (H, 3), (H, 4), (H, 5), (H, 6), (T, 1), (T, 2), (T, 3), (T, 4), (T, 5), (T, 6)}

في كثير من الأحيان يكون اهتمامك منصبا على بعض عناصر الفضاء العيني دون البعض الآخر وفي هذه الحالة تكون مهتما بما يسمى الحادث.

تعريف (9) :
الحادث Event: هو مجموعة جزئية من الفضاء العيني وبعبارة أخرى، الحادث هو مجموعة مكونة من نتيجة بسيطة واحدة أو أكثر أو لا يحتوي على أي نتائج فإذا احتوى على نتيجة بسيطة واحدة سمى حادثا بسيطا، أما إذا احتوى على نتيجتين أو أكثر فإنه يسمى حادثا مركبا.

مثال (5) :
الحادث {(H, 1)} حادث بسيط.

أما الحادث {(H, 6), (T, 2), (H, 1)} فهو حادث مركب.

وتصنف الفضاءات العينية حسب عدد العناصر التي تحتويها.

فإذا كان عدد هذه العناصر محدودا سمي الفضاء العيني فضاء عينيا محدودا. فمثلا: إذا رميت زهرة نرد فإن الفضاء العيني يحتوي على 6 نقاط: وبالتالي فإنه محدود.

تعريف (10) :

فضاء العينة المنفصل Discrete Sample Space يسمّى الفضاء العيني فضاء منفصلا إذا كان محدوداً أو لا نهائياً معدوداً، أي إذا كان محدودا أو إذا أمكن ربط عناصره واحداً إلى واحد مع الأعداد الصحيحة الموجبة، كأن تقول: أربط هذا العنصر مع العدد 1 واربط عنصرا ثانيا مع العدد 2 وأربط عنصرا ثالثا مع العدد 3، وهكذا، إلى ما لا نهاية.

مثال (6) :

إرم قطعة نقود على الأرض حتى تظهر صورة H إلى أعلى. في هذه الحالة قد ترمي قطعة النقود مرة واحدة فتحصل على صورة أو تضطر إلى رميها مرتين حيث تكون الأولى كتابة والثانية صورة فتتوقف، أو إلى رميها 3 مرات وتحصل على كتابة أولا ثم كتابة ثانيا ثم صورة في المرة الثالثة.
وهكذا، وبذلك يكون الفضاء العينيΩ = {H, TH, TTH, TTTH, ...} .

حيث H تمثل الصورة، T تمثل الكتابة. لاحظ أنك تستطيع ربط جميع عناصر Ω واحدا إلى واحد مع الأعداد الصحيحة الموجبة، فنربط (H) مع 1 و (TH) مع 2 و (TTH) مع 3 و (TTTH) مع 4 وهكذا.

تعريف (11) :

إذا كانت Ω الفضاء العيني لتجربة ما، وكان A أي حادث في Ω فإننا نعيّن لهذا الحادث عددا P(A)، بحيث يحقق الفرضيات التالية:

(1) $0 \le P(A) \le 1$

(2) $P(\Omega) = 1$

(3) إذا كان A_1 , A_2 حادثين منفصلين عن بعضهما البعض أي لا يوجد عناصر مشتركة بينهما فإن:

$$P(A_1 \cup A_2) = P(A_1) + P(A_2$$

نسمي P(A) احتمال الحادث A و P اقتران احتمال على Ω.

مثال (7) :

افرض أن Ω = {3, 2, 1}

وافرض أن ϕ هي المجموعة الخالية:

وأن

$$P(\{1\}) = P(\{2\}) = \frac{1}{4}, P(\phi) = 0$$

$$P(\{1, 2\}) = \frac{1}{2}, P(\{3\}) = \frac{1}{2}$$

$$P(\Omega) = 1, P(\{2, 3\}) = P(\{1, 3\}) = \frac{3}{4}$$

فإن P هو اقتران احتمال على Ω، وذلك لأن الفرضيات الثلاث في تعريف الاحتمال متحققة.

لاحظ أن الفرضية الأولى تعني أن احتمال أي حادث يكون كسرا غير سالب، أي أنه عدد حقيقي بين الصفر والواحد الصحيح.

أما الفرضية الثانية فتعني أن احتمال الحادث الأكيد (الفضاء العيني) يساوي واحدا. وأما الفرضية الثالثة فتعني أن احتمال اتحاد أي حادثين منفصلين عن بعضهما البعض هو حاصل جمع احتماليهما.

4 : 4 فضاء العينة ذو النقط المتساوية إمكانية الحدوث

Sample Space With Equally Likely Events

إن من أكثر الحالات التي يستعمل فيها التعريف السابق للاحتمال حالة: الفضاء العيني ذو النقط المتساوية إمكانية الحدوث كما يتضح فيما يلي :

إذا كان الفضاء العيني Ω لتجربة ما يحتوي على n من النقاط (أي أن عدد النتائج الممكنة لتجربة ما هو n) وكانت فرصة الحصول على أي نتيجة بسيطة تساوي فرصة الحصول على أية نتيجة بسيطة أخرى فإنه يقال: الفضاء العيني Ω فيه n من النقط المتساوية إمكانية الحدوث، ويكون كل حادث بسيط منفصلا عن أي حادث بسيط آخر ويكون احتمال كل حادث بسيط مساويا لاحتمال أي حادث آخر.

فإذا فرضت $\Omega = \{b_1, b_2, ... ,b_n\}$

فإن الحوادث البسيطة المتساوية في الاحتمال هي

$\{b_n\},...,\{b_2\}, \{b_1\}$

ويكون $P(\{b_1\}) = P(\{b_2\}) = ... = P(\{b_n\})$

وبما أن مجموع هذه الاحتمالات 1 فإنه يكون $P(\{b_i\}) = \frac{1}{n}$ لجميع القيم $i = 1, 2, ..., n$ ومن هذا نحصل على هذه النتيجة: إذا احتوى حادث A على عدد من النقط n(A) فإن احتمال هذا الحادث هو:

$$P(A) = \frac{n(A)}{n} = \frac{\text{عدد النقط المواتية للحادث A}}{\text{عدد نقط الفضاء العيني}}$$

أي أنه: إذا احتوى الفضاء العيني على عدد محدود من النقط وكانت هذه النقط متساوية إمكانية الحدوث فإن احتمال أي حادث في الفضاء العيني المذكور يساوي نسبة عدد النقط في ذلك الحادث إلى عدد النقط في الفضاء العيني.

مثال (8) :

رميت قطعة نقود متزنة وزهرة نرد منتظمة مرة واحدة. إن الفضاء العيني يحتوي على 12 نقطة سجلت في مثال (4). وبما أن قطعة النقود متزنة وزهرة النرد منتظمة فيمكننا اعتبار أن نقاط الفضاء العيني متساوية إمكانية الحدوث.وهذه النقاط تمثل النتائج البسيطة في التجربة وتكون احتمالاتها متساوية، إذن احتمال كل نتيجة بسيطة

$= \frac{1}{12}$

إذا كان الحادث A ={(T, 3), (H, 4), (H, 1)} فما احتمال A؟

إن عدد النقط في A هي 3

إذن $P(A) = \frac{3}{12} = \frac{1}{4}$

يتضح لك في هذا المثال إنك حصلت على عدد نقط الفضاء العيني بسهولة كما في مثال (4)، كما يتضح لك أيضا أن هناك حاجة ماسة لمعرفة عدد نقط الفضاء العيني وعدد نقط الحادث الذي تريد إيجاد احتماله، وخاصة عندما تكون نقط الفضاء العيني متساوية إمكانية الحدوث. وبناء على هذا فإنك تحتاج إلى طرق منظمة تتمكن بها من عدّ نقط الفضاء العيني والحوادث فيه، وهذا ما سندرسه في البند 4 : 6 طرق العد.

4 : 5 قوانين الاحتمال Probability Laws

يواجهنا في كثير من الأحيان تطبيقات على الاحتمالات نحتاج حيالها إلى حساب احتمالات الحوادث المركبة، ولتسهيل القيام بهذه المهمة نورد القوانين الأوليّة في الاحتمال.

نظرية (1) :

إذا رمزنا للحادث الخالي بالحرف ϕ (فاى)

وكان P الاحتمال المعرف على Ω فإن $P(\phi) = 0$

لاحظ أن الحادثϕيعني عدم حدوث أي نتيجة من النتائج الممكنة في التجربة الإحصائية، ولذلك فإنه من المعقول أن نتفق على أن $P(\phi) = 0$ علما بأنه يمكن برهنة هذه الحقيقة بسهولة باستعمال فرضيات تعريف الاحتمال.

نظرية (2) :

إذا كان A حادثا في Ω، $\overline{A}$ متممة ذلك الحادث فإن:

$$P(\overline{A}) = 1 - P(A)$$

إن نظرية (2) تعطيك المعنى التالي: إذا كان احتمال حدوث الحادث A هو P(A) فإن احتمال عدم حدوث A هو $P(\overline{A}) = 1 - P(A)$

مثال (9) :

إذا كان احتمال حصول طالب على بعثة يساوي 0.9 فما احتمال عدم حصوله على تلك البعثة؟

الحل: افترض أن الحادث A يعني حصول الطالب على بعثة، فإن الحادث $\overline{A}$ يعني عدم حصول الطالب على البعثة.

ومن النظرية (2)

$$P(\overline{A}) = 1 - P(A)$$

$$= 1 - 0.9 = 0.1$$

مثال (10) :

إذا كان احتمال وصول الطالب إلى محاضرته في الوقت المحدد يساوي 0.7 فما احتمال وصول الطالب متأخرا؟ (أي عدم وصوله في الوقت المحدّد).؟

A = الوصول في الوقت المحدد إلى المحاضرة.

$\overline{A}$ = عدم الوصول في الوقت المحدد إلى المحاضرة.

$$P(\overline{A}) = 1 - P(A)$$

$$= 1 - 0.7 = 0.3$$

نظرية (3) :

إذا كان B, A أي حادثين في الفضاء العيني Ω فإن

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

مثال (11) :

شب حريق في إحدى العمارات واتصل الحارس بمركزين من مراكز مطافئ المدينة. فإذا كان احتمال وصول الإطفائية الأولى إلى مكان الحريق خلال دقيقتين يساوي 0.9 واحتمال وصول الإطفائية الثانية خلال دقيقتين يساوي 0.8 واحتمال وصول الاثنتين معا خلال المدة نفسها يساوي 0.72 فما احتمال وصول الإطفائية الأولى أو الثانية خلال دقيقين؟

الحل: افرض A يمثل الحادث "وصول الإطفائية الأولى خلال دقيقتين".

B يمثل الحادث "وصول الإطفائية الثانية خلال دقيقتين وبالتالي يكون $A \cap B$ يمثل الحادث "وصول الاطفائيتين خلال دقيقتين".

إن احتمال وصول الإطفائية الأولى أو الثانية خلال دقيقتين هو احتمال اتحاد الحادثين B,A

وباستعمال نظرية (3)

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$= 0.9 + 0.8 - 0.72 = 0.98$$

مثال (12) :

إذا كان احتمال غياب طالب عن المحاضرة الأولى يساوي 0.20 واحتمال غيابه عن المحاضرة الثانية يساوي 0.15 واحتمال غيابه عن المحاضرتين الأولى والثانية يساوي 0.05

أ. ما احتمال غياب الطالب عن واحدة من هاتين المحاضرتين على الأقل؟

ب. ما احتمال عدم غياب الطالب عن أي من المحاضرتين؟

الحل: افرض A تمثل "الغياب عن المحاضرة الأولى"، B تمثل "الغياب عن المحاضرة الثانية".

وبذلك $A \cap B$ يمثل "الغياب عن المحاضرتين".

أ. احتمال غياب الطالب عن واحدة من المحاضرتين على الأقل يعني غيابه عن الأولى أو الثانية أو عن الاثنتين، وهذا يساوي $P(A \cup B)$

ومن نظرية (3):

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$= 0.20 + 0.15 - 0.05 = 0.3$$

ب. عدم غياب الطالب عن أي من المحاضرتين يعني أنه حضر المحاضرتين، وهذا يعني متممة الحادث "غياب الطالب عن إحدى المحاضرتين على الأقل" أي أنه متممة $A \cup B$. ومن قانون المتممة.

$$P(\overline{A \cup B}) = 1 - P(A \cup B)$$

$$= 1 - 0.3 = 0.7$$

مثال (13) :

إذا كان $P(A) = 0.5, P(B) = 0.8$ ،$P(A \cap B) = 0.4$

أوجد

أ) $P(\overline{A})$ ب) $P(\overline{B})$ جـ) $P(A \cup B)$ د) $P(\overline{A \cup B})$

الحل: أ)

$P(\bar{A}) = 1 - P(A) = 1 - 0.5 = 0.5$

ب)

$P(\bar{B}) = 1 - P(B) = 1 - 0.8 = 0.2$

جـ)

$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.5 + 0.8 - 0.4 = 0.9$

د)

$P(\overline{A \cup B}) = 1 - P(A \cup B) = 1 - 0.9 = 0.1$

نظرية (4) :

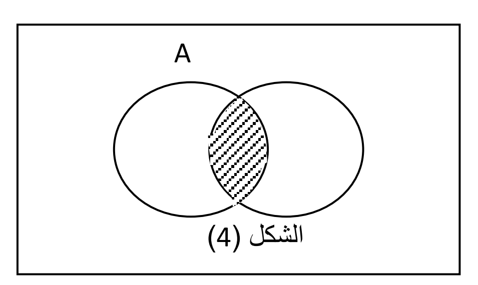
إذا كان B, A أي حادثين في فضاء العينة Ω فإن

(أ) $P(A \cap \bar{B}) = P(A) - P(A \cap B)$

(ب) وكذلك $P(B \cap \bar{A}) = P(B) - P(A \cap B)$

وهذا يعني أن احتمال حدوث A وعدم حدوث B في نفس الوقت يساوي احتمال حدوث A ناقصا احتمال حدوث الاثنين B, A معا كما يظهر في الشكل (4).

الجــزء المظلل هو $A \cap B$



الشكل (4)

أي $P(A \cap \overline{B}) = P(A) - P(A \cap B)$

لاحظ أن الفرع (أ) يمكن كتابته على الشكل $P(A)=P(A\cap B) + P(A\cap \overline{B})$ وهذا ما يسمى بقاعدة الاحتمال الكلي.

ومثل ذلك ينطبق على فرع (ب).

مثال(14) :

إذا كان احتمال حضور مدير شركة معينة في يوم ما يساوي 0.9 واحتمال حضور مساعده في ذلك اليوم 0.95 واحتمال حضور واحد منهما على الأقل يساوي 0.97 أوجد احتمال:

أ) حضور المدير ومساعده.

ب) حضور المدير وحده.

جـ) حضور مساعد المدير وحده.

الحل: أ) عبّر عن حضور المدير بالحادث A

وعن حضور المساعد بالحادث B

إذن حضور المدير ومساعده $A \cap B$

لاحظ أن حضور واحد منهما على الأقل يعبر عنه بالحادث $A \cup B$

ومن نظرية (3)

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

وبالتعويض في القانون أعلاه نجد:

$$0.97 = 0.90 + 0.95 - P(A \cap B)$$

$$0.97 = 1.85 - P(A \cap B)$$

إذن $P(A \cap B) = 1.85 - 0.97 = 0.88$

وهذا هو احتمال حضور المدير ومساعدة.

ب) حضور المدير وحده يعني أن المدير حضر ولكن مساعدة لم يحضر، ويعني ذلك حدوث A وعدم حدوث B، أي أن احتمال حضور المدير وحده يساوي احتمال حضور المدير ناقصا احتمال حضور الاثنين وذلك يعني من نظرية (4) أن

$$P(A \cap \overline{B}) = P(A) - P(A \cap B)$$

$$P(A \cap \overline{B}) = P(A) - P(A \cap B) = 0.90 - 0.88 = 0.02$$

جـ) احتمال حضور مساعد المدير وحده يعني حدوث B وليس A ويساوي

$$P(B \cap \overline{A}) = P(B) - P(A \cap B) = 0.95 - 0.88 = 0.07$$

4 : 6 طرق العد Counting Methods :

لقد درسنا في بند فضاء العينة ذي النقاط المتساوية إمكانية الحدوث أن احتمال الحادث A هو

$$P(A) = \frac{n(A)}{n}$$

حيث n عدد نقاط فضاء العينة، n(A) عدد نقاط الحادث A.

إذن، بات من الضروري أن نتعرف على طرق منتظمة لإيجاد عدد نقاط الفضاء العيني لتجربة إحصائية وعدد نقاط أي حادث في ذلك الفضاء.

(1) قاعدة الضرب Rule of Multiplication

قاعدة (1) :

إذا كانت التجربة E_1 تحدث في n_1 من الطرق وكانت التجربة E_2 تحدث في n_2 من الطرق فإن التجربتين معاً تحدثان في n_1n_2 من الطرق.

مثال (15) :

إذا أردت أن تأخذ حرفاً من المجموعة $\{a_1, a_2, a_3\}$ ثم تأخذ حرفاً من المجموعة $\{b_1, b_2\}$ لتحصل على الزوج المرتب (a_i, b_j) فإنك ستحصل على $6 = 2 \times 3$ من هذه الأزواج وهي

$(a_1, b_1), (a_1, b_2), (a_2, b_1), (a_2, b_2), (a_3, b_1), (a_3, b_2).$

ويمكن تعميم قاعدة الضرب لتشمل k من التجارب بدلا من تجربتين على النحو التالي:

قاعدة (2) :

تعميم قاعدة الضرب: إذا كانت التجربة E_1 تحدث في n_1 من الطرق، ولكل واحدة من هذه الطرق كانت التجربة E_2 تحدث في n_2 من الطرق، ولكل واحدة من هذه الطرق، كانت التجربة E_3 تحدث في n_3 من الطرق وهكذا حتى التجربة E_k التي تحدث في n_k من الطرق، فإن التجارب $E_1, E_2, \dots, E_k$ تحدث معا بعدد من الطرق يساوي $n_1, n_2, \dots, n_k$

مثال (16) :

كم خط هاتف يمكن تركيبها في مدينة، إذا تألف رقم الهاتف من سبعة أرقام أولها من اليسار 2، أو 3، أو 7

الحل: في تلك المدينة كل رقم هاتف يتألف من سبع خانات، الأولى من اليسار يوجد فيها 3 طرق فقط (لأنها تحتوي على العدد 2 أو 3 أو 7) أما باقي الخانات فتملأ كلاً منها بعشرة طرق هي الأرقام ,0, 1 9 , ... ,2 وحسب قاعدة الضرب يمكن تركيب خطوط هواتف عددها:

$$3 \times 10 \times 10 \times 10 \times 10 \times 10 \times 10 = 3000000$$

(2) قاعدة الجمع ADDITION RULE

قاعدة (3) :

إذا كانت تجربة ما تحدث في n من الطرق، وكانت تجربة ثانية تحدث في n_2 من الطرق، وكان معلوما أن التجربتين لا تحدثان معا (أي أنهما مانعتان لبعضهما البعض) فإن واحدة منهما أو الأخرى تحدث في $(n_1 + n_2)$ من الطرق.

مثال (17) :

أراد طالب شراء كتاب واحد فقط وكان لديه الاختيار من 8 كتب في الأدب و 9 كتب في الشريعة. فما عدد الاختيارات التي لديه ؟

الحل: بتطبيق قاعدة الجمع يكون لدى الطالب اختيارات عددها:

$$8 + 9 = 17$$

ويمكن تعميم قاعدة الجمع لتشمل k من التجارب على النحو التالي:

قاعدة (4):

تعميم قاعدة الجمع: إذا كان عدد التجارب k بحيث لا تحدث أي اثنتين منهما في نفس الوقت وكان عدد الطرق التي تحدث فيها الأولى n_1 وعدد طرق الثانية n_2 وهكذا، فإن عدد الطرق التي تحدث فيها واحدة من هذه التجارب هو:

$$n_1 + n_2 + ... + n_k$$

مثال (18) :

أراد طالب أن يسجل في مقرر واحد في أحد الفصول وكان لديه الاختيارات التالية:

5 مقررات إحصاء، 4 مقررات رياضيات، مقرران في الاقتصاد، 3 مقررات لغة عربية فما عدد الاختيارات التي لديه لتسجيل مقرر واحد فقط؟

الحل: حسب قاعدة الجمع يكون عدد الاختيارات

$5 + 4 + 2 + 3 = 14$

(3) التباديل Permutations

التباديل هي طرق ترتيب جميع أو بعض عناصر مجموعة ما.

مثال (19) :

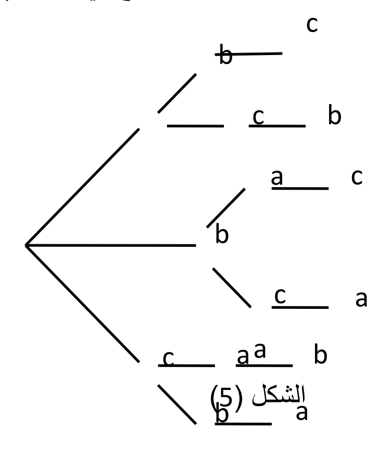
ما عدد تباديل حروف المجموعة {c, b, a}

الحل : عدد التباديل هو عدد الطرق التي يتم فيها ترتيب الحروف c, b, a.

لديك ثلاثة أماكن لتملأها ولديك ثلاثة حروف يمكن اختيار حرف واحد منها لملء المكان الأول ولديك حرفان لملء المكان الثاني ويبقى حرف واحد لملء المكان الثالث، وبتطبيق قاعدة الضرب يكون عدد الطرق

$3 \times 2 \times 1 = 6$

ويمكن تمثيل التباديل الستة السابقة على شكل شجرة كما هو موضح في الشكل (5).



الشكل (5)

وبوجه عام يمكنك ترتيب n من العناصر المميزة بطرق عددها

a

$n(n - 1)(n - 2) .... 3. 2. 1$

والعادة أن يستعمل الرمز !n الذي يقرأ "مضروب n" ليدل على حاصل الضرب السابق ويستعمل الرمز nPn ليدل على عدد تباديل n من العناصر المميزة فيكون

$$nPn = n! = n(n - 1)(n - 2) .... 3. 2. 1$$

وهذا هو عدد تباديل n من العناصر المميزة.

مثال (20) :

بكم طريقة يمكن ترتيب حروف كلمة "تقوى"؟

الحل: هناك 4 حروف مختلفة عن بعضها البعض (مميزة) هي ت، ق، و، ى وحسب تعريف التباديل يكون عدد طرق الترتيب.

$4! = 4 \times 3 \times 2 \times 1 = 24$

أما إذا أردت أن ترتب بعض عناصر مجموعة فهذا يعني أنك لا تأخذ العناصر وترتبها في أماكن عددها مساوٍ لعدد العناصر ولكن يكون لديك عدد من الأماكن أقل من عدد العناصر، وهذا ما نسميه بعدد تباديل n من العناصر المميزة إذا أخذت r في كل مرّة، ونجد عدد هذه التباديل كما يلي:

إذا كان لديك n من العناصر المميزة ولديك r من الأماكن في صف فبكم طريقة ترتب العناصر في هذه الأماكن بحيث تضع عنصراً واحداً في كل مكان. يمكن أن تنظر للمسألة كالتالي:

لملء المكان الأول لديك n من الاختيارات.

لملء المكان الثاني لديك (n - 1) من الاختيارات.

لملء المكان الثالث لديك (n - 2) من الاختيارات.

وهكذا إلى أن تصل إلى المكان الأخير الذي رقمه r فيكون لديك (n - r + 1) من الاختيارات لملء هذا المكان. وبتطبيق قاعدة الضرب يكون عدد التباديل

$nPr = n(n - 1)(n - 2) .... (3. 2. 1)$

وإذا ضربت هذا العدد بالمضروب !(n - 1) وقسمته على نفس ذلك المضروب، تحصل على: تباديل n من العناصر المميزة مأخوذة r في كل مرّة يساوي:

$$nPr = n(n-1)(n-2) \ldots 3.\,2.\,1 \,.\, \frac{(n-r)!}{(n-r)!}$$

$$= \frac{n!}{(n-r)!}$$

مثال (21) :

كم كلمة ذات ثلاثة حروف مختلفة يمكنك صياغتها من الحروف أ، ب، جـ، د، هـ. حتى ولو لم يكن للكلمة معنى لغوي.

الحل: المطلوب هو بكم طريقة يمكنك ترتيب 5 حروف مميزة (مختلفة) مأخوذة 3 في كل مرّة. أي عدد تباديل 5 حروف مأخوذة 3 في كل مرّة، وهذا يساوي:

$${}_5P_3 = 5.4.3 = 60$$

لأنه يمكنك ملء المكان الأول بواحد من الحروف الخمسة والمكان الثاني بواحد من الحروف الأربعة المتبقية والمكان الثالث بواحد من الحروف الثلاثة المتبقية وباستعمال قاعدة الضرب نحصل على الجواب.

وبعبارة أخرى، عدد تباديل 5 حروف مميزة مأخوذة 3 في كل مرّة هو

$${}_5P_3 = \frac{5!}{(5-3)!} = \frac{5!}{2!} = \frac{5.4.3.2.1}{2.1} = 60$$

(4) التوافيق Combinations

التوافيق هي الطرق التي نختار بها عدداً معيناً من عناصر مجموعة معينة دون النظر إلى الترتيب، وبعبارة أخرى هي عدد الطرق التي نقسم بها مجموعة من العناصر إلى مجموعتين تحتوي إحداهما عدداً معيناً من العناصر وتحتوي الأخرى بقية العناصر دون النظر إلى ترتيب تلك العناصر في أي من المجموعتين.

مثال (22)

ما عدد الطرق التي تختار بها حرفين من الحروف أ، ب، جـ دون الاهتمام بالترتيب؟

الحل: الاختيارات هي }أ، ب{، }أ، جـ{، }ب، جـ{ وبذلك يكون عدد الطرق = 3 وبوجه عام، ما عدد الطرق التي تختار بها r عنصراً من مجموعة فيها n من العناصر دون النظر إلى الترتيب.

إن عدد الطرق المطلوب هو ما يدعى عدد توافيق n من العناصر مأخوذة r في كل مرة. فإذا عبّرنا عن عدد التوافيق هذه بالرمز $\binom{n}{r}$ (وتقرأ n فوق r) نحصل على القاعدة:

قاعدة (5)

عدد الطرق التي نختار بها r عنصرا من مجموعة فيها n من العناصر دون النظر إلى الترتيب هو عدد توافيق n من العناصر مأخوذة r في كل مرة ويساوي :

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

مثال (23) :

صف فيه 10 طلاب، بكم طريقة تختار لجنة مؤلفة من 3 طلاب دون وجود مناصب لأعضاء اللجنة؟

الحل: **عدد الطرق المطلوب، حسب قاعدة التوافيق هو:**

$$\binom{10}{3} = \frac{10!}{3!(10-3)!}$$

$$\frac{10!}{3!7!} = \frac{10.9.8.7.6.5.4.3.2.1}{3.2.1.7.6.5.4.3.2.1} = 120$$

مثال (24) :

صندوق فيه 5 كرات حمراء و 7 كرات بيضاء.

(أ) بكم طريقة تختار 4 كرات من الصندوق؟

(ب) بكم طريقة تختار الكرات الأربع بحيث تكون فيها كرة واحدة حمراء وثلاث كرات بيضاء؟

(جـ) ما احتمال الحصول على كرة حمراء وثلاث كرات بيضاء عند اختيار 4 كرات من الصندوق؟

الحل:

(أ) من قاعدة التوافيق، عدد طرق اختيار 4 كرات من الصندوق = $\binom{12}{4}$

وذلك لأن مجموع الكرات 12 ونريد أن نختار منها 4 فيكون عدد طرق الاختيار:

$$\binom{12}{4} = \frac{12!}{4!(12-4)!} = \frac{12!}{4!8!} = \frac{12.11.10.9.8!}{4!8!}$$

$$= \frac{12.11.10.9}{4.3.2.1} = 495$$

(ب) عدد طرق اختيار كرة واحدة من الكرات الحمراء هو

$5 = \frac{5!}{1!4!} = \binom{5}{1}$ لأنه لدينا 5 كرات حمراء ونريد اختيار واحدة منها فقط.

عدد طرق اختيار 3 كرات بيضاء هو

$35 = \frac{7.6.5.4!}{3.2.4!} = \frac{7!}{3!4!} = \binom{7}{3}$ لأنه لدينا 7 كرات بيضاء ونريد اختيار 3 منها.

إذن، من قاعدة الضرب، عدد طرق اختيار كرة واحدة حمراء وثلاث كرات بيضاء هو

$$5 \times 35 = 175$$

(جـ) عدد النقاط في الفضاء العيني هو 495 من فرع (أ)،

ومن فرع (ب) عدد نقاط الحادث المطلوب هو 175

إذن، الاحتمال المطلوب هو $\frac{175}{495}$

ويمكن كتابة جواب (جـ) مباشرة:

احتمال الحصول على كرة حمراء وثلاث كرات بيضاء عند اختيار 4 كرات من الصندوق هو:

$$P(\text{1 حمراء، 3 بيضاء}) = \frac{\binom{5}{1}\binom{7}{3}}{\binom{12}{4}} = \frac{175}{495}$$

= 0.354

مثال (25) :

لدى مستودع الجامعة 12 حاسبة إلكترونية منها آلتان عاطلتان. تسلمت إحدى الدوائر 4 آلات اختيرت عشوائيا من هذا المستودع.

(أ) ما احتمال عدم وجود أي آلة عاطلة ضمن ما استلمته الدائرة؟

(ب) ما احتمال وجود آلة عاطلة واحدة ضمن ما استلمته الدائرة؟

الحل:

(أ) احتمال عدم وجود آلة عاطلة ضمن ما استلمته الدائرة يساوي احتمال 4 آلات صالحة نختارها من 8 آلات صالحة،

$$\text{P(4 صالحة، 0 عاطلة)} = \frac{\binom{4}{0}\binom{8}{4}}{\binom{12}{4}} = \frac{8!/4!.4!}{12!/4!8!}$$

$$= \frac{70}{495} = 0.141$$

مثال (26) :

ثلاثة طلاب في غرفة، ما احتمال أن لا يكون من بينهم أي اثنين أو أكثر لهم تاريخ الميلاد نفسه؟ أي ما احتمال أن تكون تواريخ ميلادهم مختلفة عن بعضها البعض؟

الحل: أي نقطة في الفضاء العيني هي عبارة عن ثلاثة أعداد مرتبة، العدد الأول هو تاريخ ميلاد الطالب الأول، العدد الثاني هو تاريخ ميلاد الطالب الثاني والعدد الثالث هو تاريخ ميلاد الطالب الثالث. وبما أن تاريخ ميلاد أي طالب هو أحد أيام السنة أي أنه يوجد 365 يوما يمكن أن يكون تاريخ ميلاد الطالب واحدا منها.

إذاً يكون عدد نقاط الفضاء العيني $(365)^3$ ويكون 365 × 364 × 363 هو عدد النقاط المواتية للحادث المطلوب وذلك لأن الطالب الأول أمامه 365 يوماً ليكون تاريخ ميلاده فيها، ولكن الطالب الثاني يوجد أمامه 364 فقط، لأنه لا يكون تاريخ ميلاده نفس تاريخ ميلاد الطالب الأول، وهكذا يوجد للطالب الثالث 363 يوما ليكون تاريخ ميلاده فيها وذلك لكي تكون تواريخ ميلاد الطلاب الثلاثة مختلفة عن بعضها البعض.

إذاً الاحتمال المطلوب هو = $\frac{365 \times 364 \times 363}{(365)^3}$

لقد وجدنا في القاعدة (5) عدد الطرق التي نختار بها r عنصراً من مجموعة فيها n عنصراً، أي عدد الطرق التي نقسم بها مجموعة فيها n من العناصر إلى مجموعتين واحدة فيها r من العناصر والثانية فيها باقي العناصر وعددها (n - r).

والآن إذا أردنا تقسيم مجموعة مؤلفة من n عنصراً إلى k من المجموعات الجزئية، الأولى فيها n_1 عنصراً والثانية n_2 عنصراً، وهكذا، حتى نصل إلى المجموعة التي رقمها k وفيها n_k عنصراً، فبكم طريقة نحصل على هذا التقسيم؟

إن الإجابة هي تعميم القاعدة (5) وهي:

القاعدة (6) :
عدد الطرق التي نقسم بها مجموعة فيها n من العناصر إلى k مجموعات جزئية بحيث المجموعة الأولى تحوي n_1 عنصراً والمجموعة الثانية تحوي n_2 عنصراً، وهكذا، حتى نصل إلى المجموعة k وتحوي n_k عنصراً هو

$$\frac{n!}{n_1!n_2!...n_k!}$$

مع ملاحظة أن $n_1 + n_2 + ... + n_k = n$

مثال (27) :
بكم طريقة يمكنك أن تقسم 30 طالباً إلى أربع مجموعات تحوي على التوالي، 7، 10، 5، 8 طلاب.

الحل: $\frac{30!}{7!10!5!8!}$ بالتطبيق المباشر للقاعدة (6).

لاحظ أن بالإمكان استعمال القاعدة (6) في إيجاد عدد تباديل مجموعة من العناصر إذا كان فيها عناصر متماثلة من نوع واحد وعناصر أخرى متماثلة من نوع آخر وهكذا.

قاعدة (7) :
عدد تباديل مجموعة فيها n من العناصر، إذا كان ضمنها n_1 من العناصر المتماثلة، n_2 من العناصر المتماثلة المختلفة عن الأولى، .... ، n_k من العناصر المتماثلة المختلفة عن سابقاتها هو :

$$\frac{n!}{n_1!n_2!...n_k!}$$

حيث $n_1 + n_2 + ... + n_k = n$

مثال (28) :

بكم طريقة ترتب حروف كلمة: ميسيسبي

الحل: الكلمة فيها 7 حروف، منها 1م، 3 ي، 2 س، 1 ب.

إذاً بتطبيق القاعدة (7)، عدد الطرق هو

$$\frac{7!}{1!3!2!1!} = \frac{7!}{3!2!} = 420$$

مثال (29) :

ما عدد تباديل حروف كلمة STATISTICS

الحل: العدد يساوي $\frac{10!}{3!3!2!1!1!}$

4 : 7 الاحتمال الشرطي Conditional Probability

إن دراسة الأحداث المشروطة تعني دراسة حدوث حادث معين إذا علم تحقق حدوث حادث آخر، وهذه دراسة تستلزم دراسة الاحتمال الشرطي الذي يعني دراسة احتمال حادث ما إذا علم حدوث حادث آخر. فمثلا إذا رميت زهرة نرد منتظمة وعلمت أن العدد على الوجه الظاهر كان زوجيا، فما احتمال أن يكون العدد على الوجه الظاهر هو 2؟

من الواضح أنه إذا علمت أن العدد الظاهر على الوجه زوجي فهذا يعني أنك علمت حدوث ظهور العدد الزوجي وهذا يعني أن العدد الذي ظهر يجب أن يكون من المجموعة {2, 4, 6}.

الآن ما احتمال أن يكون العدد الظاهر 2؟

بما أن زهرة النرد منتظمة وعلمنا حدوث {2, 4, 6} فاحتمال ظهور 2 يساوي $\frac{1}{3}$.

فإذا عبّرت بالرمز E للحادث "العدد الظاهر زوجي".

فيكون المطلوب احتمال ظهور العدد 2 إذا علمت أن العدد الظاهر زوجي.

هذا هو الاحتمال الشرطي، ونعبر عنه بالرمز $P(\{2\} | E)$

تعريف (12) :

الاحتمال الشرطي للحادث A إذا علم الحادث B هو

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

شريطة أن $P(B) > 0$.

مثال (30) :

رميت قطعة نقود ثلاث مرات فإذا رمزنا لظهور الصورة بالحرف H. وظهور الكتابة بالحرف T، وإذا علم أن الوجه في الرمية الأولى H فما احتمال أن يكون الوجهان الآخران H , H؟

الحل: الفضاء العيني هو:

$$\Omega = \begin{Bmatrix} H & H & H \\ H & H & T \\ H & T & H \\ H & T & T \\ T & H & H \\ T & H & T \\ T & T & H \\ T & T & T \end{Bmatrix}$$

إفرض A تمثل الحادث "الوجه في الرمية الأولى H".

B تمثل الحادث "الوجهان في الرميتين الثانية والثالثة H, H".

والمطلوب $P(B \mid A)$

من التعريف $P(B \mid A) = \frac{P(B \cap A)}{P(A)}$

بالنظر إلى فضاء العينة Ω فإن

$P(B \cap A) = \frac{1}{8}$

وذلك لأن في $A \cap B$ نقطة واحدة HHH وفي الفضاء العيني 8 نقاط.

وكذلك $P(A)=\frac{4}{8}$ لأن الحادث{A = {HHH, HHT, HTH, HTT أي فيه 4 نقاط

إذن = $P(B \mid A) = \frac{1}{4} = \frac{1/8}{4/8}$

بالرجوع إلى التعريف السابق وبعد إجراء الضرب التبادلي نجد :

$$P(A \cap B) = P(B) . P(A \mid B)$$

وكذلك $P(B \cap A) = P(A) . P(B \mid A)$

ولكن $P(A \cap B) = P(B \cap A)$ لأن العناصر المشتركة بين B, A هي نفسها العناصر المشتركة بين A, B ولذلك نحصل على النظرية :

نظرية (5) :

قاعدة الضرب Multiplication Rule

$$\mathbf{P(A \cap B) = P(A) . P(B \mid A)}$$

إذا كان $P(A) > 0$

أو $P(A \cap B) = P(B) . P(A \mid B)$

إذا كان $P(B) > 0$

وهذه النظرية تعطيك قاعدة الضرب في حالة الأحداث المشروطة.

مثال (31):

إذا كان $P(B) = 0.3, P(A) = 0.6$ وكان $P(A \mid B) = 0.4$

أوجد

أ) $P(A \cap B)$ ب) $P(B \mid A)$

الحل:

أ) من قاعدة الضرب

$$P(A \cap B) = P(B) . P(A \mid B)$$

$$= 0.3 \times 0.4 = .12$$

ب) من تعريف الاحتمال الشرطي:

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{0.12}{0.6}$$

$$= 0.2$$

مثال (32) :

صندوق فيه 3 كرات حمراء، 9 كرات بيضاء، سحبت منه كرة واحدة وشوهد لونها وطرحت جانبا ثم سحبت منه كرة أخرى. أوجد احتمال أن "الكرة الأولى بيضاء والثانية حمراء"

الحل: افرض أن الحادث A يمثل "الكرة الأولى بيضاء".

وافرض أن الحادث B يمثل "الكرة الثانية حمراء".

إذا فالمطلوب $P(A \cap B)$

ومن قاعدة الضرب: $P(A \cap B) = P(A) . P(B \mid A)$

ولكن، بما أن عدد الكرات كلها 12 وعدد الكرات البيضاء 9 ولذلك يكون $P(A) = \frac{9}{12}$ وأيضا $P(B \mid A) = \frac{3}{11}$ لأن احتمال الكرة الثانية حمراء إذا علم أن الكرة الأولى بيضاء يعني أنه بعد أن علمت أن الأولى بيضاء وطرحتها جانبا فإنه قد بقي في الصندوق 3 كرات حمراء و 8 كرات بيضاء.

وبالتالي يكون= $P(B \mid A)$ $\frac{3}{11}$

إذن، ومن قاعدة الضرب تجد:

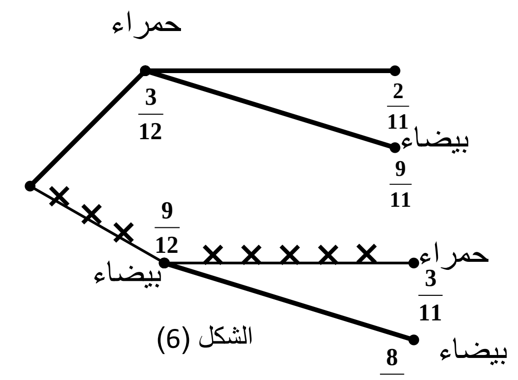
$$P(A \cap B) = \frac{9}{12} \times \frac{3}{11} = \frac{9}{44}$$

ويمكن توضيح قاعدة الضرب باستعمال شجرة الاحتمال حيث نمثل الفضاء العيني بأصل الشجرة ونمثل الحوادث بفروع الشجرة.

ثم نمثل الحوادث الشرطية بفروع جديدة لكل فرع وهكذا. في هذه الحال تكون احتمالات الفروع الأولى احتمالات عادية أما احتمالات الفروع الثانية فتكون احتمالات شرطية، حيث علم تحقق حدوث الفروع التي قبلها. وبضرب الاحتمالات العادية بالاحتمالات الشرطية نحصل على قاعدة الضرب. لتوضيح هذه الفكرة نحل لك المثال (32) ثانية.

من أصل شجرة الاحتمال يتفرع فرعان، واحد يعطي كرة حمراء واحتماله $\frac{3}{12}$ والثاني يعطي كرة بيضاء واحتماله $\frac{9}{12}$.

ومن الفرع الأول يتفرع فرعان جديدان، واحد يعطي كرة حمراء واحتماله (احتمال شرطي) $\frac{2}{11}$ والثاني يعطي كرة بيضاء واحتماله (احتمال شرطي) $\frac{9}{11}$ وكذلك من الفرع الثاني يتفرع فرعان جديدان يعطي الأول كرة حمراء واحتماله $\frac{3}{11}$ ويعطي الثاني كرة بيضاء واحتماله $\frac{8}{11}$ كما يظهر في الشكل (6).



الشكل (6)

إن احتمال "الكرة الأولى بيضاء والثانية حمراء" هو احتمال بيضاء ثم حمراء الذي يتمثل على الخط المشار عليه بالنقاط ××× وهو حاصل الضرب.

$$\frac{9}{12} \times \frac{3}{11} = \frac{9}{44}$$

4 : 8 الحوادث المستقلة Independent Events

ربما تسأل نفسك: هل يؤثر علمك بحدوث حادث A في احتمال حدوث حادث آخر B؟

فمثلاً، إذا رميت قطعة نقود وظهرت صورة فهل تؤثر هذه الحقيقة في احتمال ظهور العدد 5 عند رمي زهرة نرد منتظمة؟

في كثير من الحالات لا يتغير احتمال حدوث حادث B ولا يتأثر بعلمك أن حادثا A قد حدث، أي أن: احتمال حدوث B إذا علمت أن A قد حدث، يبقى مساويا لاحتمال حدوث B في الأصل، أي أن $P(B \mid A) = P(B)$

في هذه الحالة نقول إن الحادث B مستقل عن الحادث A. وبالمثل نقول إن الحادث A مستقل عن الحادث B إذا كان

$$P(A \mid B) = P(A)$$

مما سبق يظهر لك أنه إذا كان B مستقلا عن A فإن

$P(B \mid A) = P(B)$ وبالتعويض في نظرية (5) تجد:

$$P(A \cap B) = P(A) \,.\, P(B)$$

وبالمثل إذا كان A مستقلا عن B فإن $P(A \mid B) = P(A)$ وبالتالي نستعمل التعريف التالي:

تعريف (13) :

يكون الحادثان B, A مستقلين إذا كان: $P(A \cap B) = P(A) \,.\, P(B)$

والعكس صحيح.

إن التعريف (13) يعطيك قانون ضرب الاحتمالات عندما تكون الحوادث مستقلة.

مثال (33) : إذا كان B, A مستقلين وكان

$P(B) = 0.6, P(A) = 0.4$ أوجد $P(A \cap B)$

الحل: بما أن الحادثين مستقلان، إذن

$$P(A \cap B) = P(A) . P(B)$$
$$= 0.4 \times 0.6 = 0.24$$

مثال (34) :

في مدينة ما اطفائيتان تعملان مستقلتين عن بعضهما البعض، احتمال وصول الأولى إلى مكان حريق خلال خمس دقائق 0.95 واحتمال وصول الثانية إلى المكان خلال المدة نفسها يساوي 0.90، ما احتمال وصول الاطفائيتين إلى مكان الحريق خلال خمس دقائق.

الحل: افرض A يمثل وصول الاطفائية الأولى خلال خمس دقائق، وأن B يمثل وصول الاطفائية الثانية خلال خمس دقائق، وبما أن الاطفائيتين مستقلتان عن بعضهما البعض.

إذن احتمال وصول الاطفائيتين إلى مكان الحريق يساوي :

$$P(A \cap B) = P(A) . P(B)$$
$$= 0.95 \times 0.90 = 0.855$$

مثال (35) :

إذا كان $P(A) = 0.6$، $P(A| B) = 0.6$، $P(B| A) = 0.5$

هل B, A مستقلان؟ وما احتمال B

الحل: بما أن $P(A| B) = 0.6$ وأن $P(A) = 0.6$

إذا $P(A| B) = P(A)$

وبالتالي فإن B, A مستقلان.

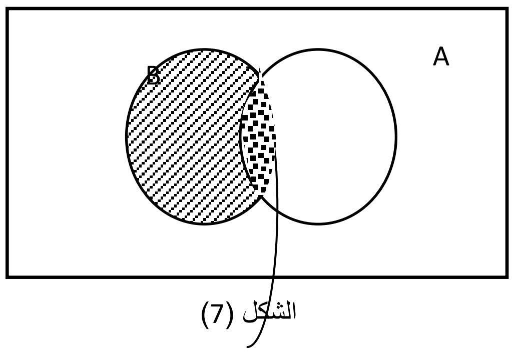
والآن، بما أن B, A مستقلان فإن:

$$P(B) = P(B| A) = 0.5$$

وبالتالي يكون $P(B) = 0.5$

4 : 9 نظرية بييز Bayes' Theorem

إذا كان Ω الفضاء العيني لتجربة ما، وكان A حادثا في Ω بحيث $P(A) > 0$ وكان B أي حادث في Ω كما يظهر في الشكل (7) حاول أن تحسب احتمال B



الشكل (7)

من نظرية (4) فرع (ب) نجد أن

$$P(B) = P(B \cap A) + P(B \cap \overline{A})$$

ومن قانون الضرب للحوادث المشروطة تجد:

$$P(B) = P(A) \,.\, P(B| A) + P(\overline{A})\, P(B| \overline{A})$$

وهذا ما يدعى بقانون الاحتمال الكلي.

وبالاستفادة من المعادلة السابقة يمكنك حساب $P(A| B)$ حيث من تعريف الاحتمال الشرطي.

$$P(A| B) = \frac{P(A \cap B)}{P(B)}$$

ولكن من قانون الضرب للحوادث المشروطة

$$P(A \cap B) = P(A)P(B| A)$$

وبتعويض قيمة $P(B)$ من معادلة الاحتمال الكلي تجد:

$$P(A| B) = \frac{P(A).P(B | A)}{P(A)P(B | A) + P(\overline{A}).P(B | \overline{A})}$$

وهذه قاعدة بييز في حالة حادثين.

أما في الحالة العامة فنحتاج إلى التعريف التالي:

تعريف (13) :

نقول إن $A_1, A_2, ..., A_n$ تجزئة للفضاء العيني Ω إذا كان:

(أ) جميع الحوادث $A_1, A_2, ..., A_n$ منفصلة ثنائيا أي أن $A_i \cap A_j = \phi$ لكل $i \neq j$ حيث i من 1 إلى n و j من 1 إلى n.

(ب) اتحاد $A_1, A_2, ..., A_n$ يساوي الفضاء العيني، أي أن:

$$A_1 \cup A_2 \cup .... \cup A_n = \Omega$$

نظرية بييز :

إذا كانت $A_1, A_2, ..., A_n$ تجزئة للفضاء العيني Ω وكان $P(A_i) > 0$ لكل i من 1 إلى n.

وكان E أي حادث في Ω بحيث $P(E) > 0$ فإن:

$$P(A_k|E)=\frac{P(A_k)P(E|A_k)}{P(A_1)P(E|A_1)+P(A_2)P(E|A_2)+...+P(A_n)P(E|A_n)}$$

حيث k أي عدد صحيح من 1 إلى n.

وإذا أردنا كتابة نظرية بييز في حال ثلاثة حوادث، فهي كما يلي:

إذا كان A_3, A_2, A_1 تجزئة للفضاء العيني Ω وكان $P(A_i) > 0$ لجميع $i = 1, 2, 3$ وكان E أي حادث في Ω بحيث $P(E) > 0$ فإن

$$P(A_1| E) = \frac{P(A_1).P(E|A_1)}{P(A_1)P(E|A_1)+P(A_2)P(E|A_2)+P(A_3)P(E|A_3)}$$

وبنفس الطريقة تجد $P(A_2| E)$، وذلك بأن تضع في البسط $P(A_2).P(E| A_2)$. وبالمثل تجد $P(A_3| E)$ وذلك بأن تضع في البسط $P(A_3). P(E| A_3)$.

وكما تم استعمال شجرة الاحتمال لتفسير قاعدة الضرب في حالة الاحتمال الشرطي فإنه يمكن استعمال تلك الشجرة في حساب الاحتمال الكلي وتفسير نظرية بييز، وذلك باعتبار أصل الشجرة الفضاء العيني،

P(A1)
P(E|A1)
P(A2)
P(E|A2)
E

والفروع الأولى منها تمثل تجزئة الفضاء العيني، ثم يتفرع كل فرع إلى فروع ثانوية تؤدي إلى حادث معين. تكون الاحتمالات في الفروع الأولى (تجزئة الفضاء العيني) احتمالات عادية ومجموعها 1، وفي الفروع الثانوية احتمالات شرطية. ويتضح ذلك من الشكل (8) حيث أن تجزئة الفضاء العيني كانت A_1, A_2, A_3 والحادث المطلوب حساب احتماله، E

الشكل (8)

من الواضح أننا نحصل على P(E) من الفروع الثلاثة ولذلك نجد:

$$P(E) = P(A_1) .P(E| A_1) + P(A_2) .P(E| A_2) + P(A_3) .P(E| A_3)$$

أما نظرية بييز فنحصل عليها كما يلي:

$$P(A_1| E) = \frac{P(A_1).P(E|A_1)}{P(A_1)P(E|A_1)+P(A_2)P(E|A_2)+P(A_3)P(E|A_3)}$$

وبنفس الطريقة تحصل على $P(A_3| E), P(A_2| E)$

مثال (36) :

مصنع فيه ثلاث ماكنات، تصنع كل منها $\frac{1}{3}$ ناتج المصنع، فإذا كان احتمال التالف من إنتاج الماكنة الأولى يساوي 0.03 ومن إنتاج الثانية 0.02 ومن إنتاج الثالثة 0.04، أخذنا عنصرا واحدا من ناتج المصنع فوجدناه تالفا

(أ) ما احتمال أنه صنع بواسطة الماكنة الأولى؟

(ب) ما احتمال أنه صنع بواسطة الماكنة الثالثة؟

الحل: افرض الماكنة الأولى A_1 والثانية A_2 والثالثة A_3

فيكون $P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}$

افرض E تمثل "العنصر التالف"، فيكون

$P(E \mid A_1) = 0.03, P(E \mid A_2) = 0.02, P(E \mid A_3) = 0.04$

(أ) المطلوب هو $P(A_1 \mid E)$

بتطبيق نظرية بييز.

$$P(A_1|E) = \frac{P(A_1).P(E \mid A_1)}{P(A_1).P(E \mid A_1) + P(A_2).P(E \mid A_2) + P(A_3)P(E \mid A_3)}$$

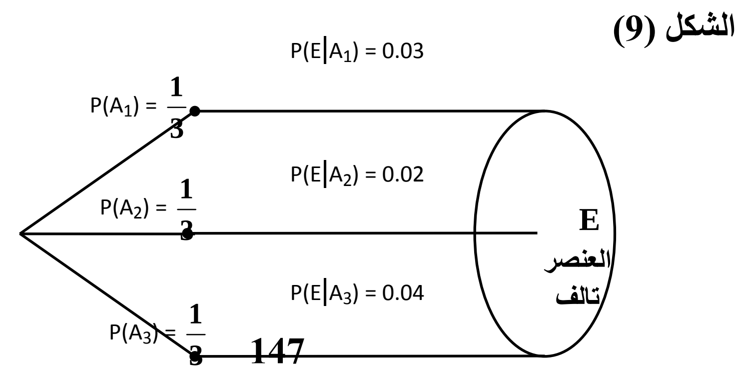
$$= \frac{\frac{1}{3} \times 0.03}{\frac{1}{3} \times 0.03 + \frac{1}{3} \times 0.02 + \frac{1}{3} \times 0.04}$$

$$= \frac{0.01}{0.03} = \frac{1}{3}$$

ب) بنفس الطريقة

$$P(A_3 \mid E) = \frac{\frac{1}{3} \times 0.04}{\frac{1}{3} \times 0.03 + \frac{1}{3} \times 0.02 + \frac{1}{3} \times 0.04} = \frac{0.04}{0.09} = \frac{4}{9}$$

ويمكن الاستعانة بشجرة الاحتمال في حل المثال (36) كما يظهر في الشكل (9)



الشكل (9)

تمارين

1-4 : إذا كانت الحوادث A, B, C, D منفصلة ثنائياً وكان اتحادها يساوي فضاء العينة Ω، أذكر أي التعيينات التالية تمثل (تحقق شروط) الاحتمال وأيها لا تفي بالغرض مع ذكر السبب.

(أ) $P(A) = P(B) = \frac{1}{4}, P(C) = \frac{1}{6}, P(D) = \frac{1}{3}$

(ب) $(A) = \frac{1}{2}, P(B) = \frac{1}{3}, P(C) = \frac{1}{4}, P(D) = 0$

(جـ) $P(A) = P(B) = P(C) = \frac{1}{4}, P(D) = \frac{1}{3}$

(د) $P(C) = -0.1, P(D) = 0.2, P(A) = 0.7, P(B) = 0.2$

(هـ) $P(A) = P(C) = P(D) = \frac{1}{4}, P(B) = 0$

2-4 : إذا كان $P(A) = 0.6 , P(B) = 0.3,$

$P(A \cap B) = 0.2$

أوجد $P(A \cup B), P(\overline{A} \cap B), P(\overline{A} \cup \overline{B})$

3-4 : عرّف فضاء العينة Ω لكل تجربة مما يأتي:

(أ) أعطيت خمسة أصناف من الشاي من الأنواع A, B, C, D, E لخبراء في المذاق وطلب منهم أن يسجلوا النوع ذا المذاق الأفضل والنوع ذا المذاق الأسوأ.

كم نقطة في فضاء العينة الذي حصلت عليه. اكتب نقاط فضاء العينة.

(ب) ارم قطعة نقود 4 مرات ولاحظ ظهور الصورة أو الكتابة في كل رمية.

4-4 : باستعمال شكل فن (Venn) تحقق مما يلي:

a) $(\overline{A \cap B}) = \overline{A} \cup \overline{B}$

b) $(A \cap B) \cup (A \cap C) = A \cap (B \cup C)$

c) $\overline{A} \cap \overline{B} \cap \overline{C} = (\overline{A \cup B \cup C})$

5-4 : يشعر السيد أحمد أن احتمال فوز فريق الوحدات في المباراة الأولى $\frac{2}{3}$ وفوزه في المباراة الثانية $\frac{5}{12}$ وفوزه في المبارتين معاً $\frac{1}{24}$ ، فهل يتفق هذا الشعور مع تعريف الاحتمال؟ اشرح السبب

6-4 : إذا كان $P(A) = 0.4 , P(B) = 0.5 , P(C) = 0.6 , P(A \cap C) = 0.2$

(أ) إذا كان A , B مستقلين، احسب : $P(\overline{A} \cap \overline{B})$.

(ب) إذا كان A , B منفصلين، احسب $P(B \cap \overline{A}), P(\overline{A} \cap \overline{B})$.

(جـ) احسب $P(\overline{A} | C), P(A | \overline{C})$

(د) هل A ,C مستقلان ؟

7-4 : سحبت 5 أوراق من مجموعة ورق اللعب (52 ورقة).

(أ) ما احتمال أن تكون ورقتان من اللون الأسود وثلاث أوراق من اللون الأحمر؟

(ب) ما احتمال أن يكون ضمن الأوراق الخمسة اثنتان من لون وثلاثة من لون آخر ؟

(جـ) ما احتمال أن تتألف الأوراق الخمسة من عشرتين وخمستين وولد ؟

(د) ما احتمال أن تتألف الأوراق الخمسة من صورة، 7، 5، 3، 1 ؟

8-4 : يوجد في جهاز أداتان أ ، ب. فإذا كان احتمال فشل (أ) يساوي 0.02 وفشل (ب) يساوي 0.03 وفشل الاثنتين معا 0.01.

(i) هل الحادثان فشل أ وفشل ب مستقلان ؟

(ii) ما احتمال فشل (ب) إذا علم أن (أ) قد فشلت ؟

(iii) ما احتمال فشل واحدة من الأداتين بالضبط ؟

4-9 : يعمل ثلاثة عمال أ، ب، جـ في مصنع. فإذا كانت نسبة ما ينتجه أ هي 30% من الناتج الكلي ونسبة ب هي 50% ونسبة جـ هي 20%، وإذا كان 5% من ناتج أ و 3% من ناتج ب، 4% من ناتج جـ معيباً، اخترنا أداة من ناتج هذا المصنع.

(i) ما احتمال أن تكون هذه الأداة معيبة ؟

(ii) إذا كانت الأداة التي اخترناها معيبة فما احتمال أنها من إنتاج العامل (ب)؟

4-10 : كيس فيه 4 كرات بيضاء، 3 كرات خضراء وكلها متشابهة في الحجم والملمس. اخترنا كرتين من هذا الكيس.

(أ) ما احتمال أن الكرتين خضراوان ؟

(ب) ما احتمال أن احدى الكرتين بيضاء والأخرى خضراء ؟

4-11 : بكم طريقة يمكن ترتيب حروف كلمة MISSISSIPPI ؟

4-12 : (أ) بكم طريقة يمكن أن يجلس 5 طلاب في صف ؟

(ب) بكم طريقة يمكن أن يجلس 5 طلاب حول طاولة مستديرة ؟

4-13 : رتبت حروف كلمة MATHEMATICS ، ما احتمال الحصول على كلمة تبدأ ب MATH ؟

4-14 : صف فيه 10 طلاب (أ) بكم طريقة نختار لجنة مكونة من 3 طلاب ؟

(ب) بكم طريقة نختار لجنة مكونة من 3 طلاب على أن يكون أحدهم رئيسا للجنة ؟

4-15 : لدى عائلة 4 أطفال ، ما احتمال أن يكون جميعهم ذكورا ؟

4-16 : بكم طريقة يمكن ترتيب حروف كل من الأسماء : "لمى"، "طارق" ، "عبد القادر"، "محمّد".

الفصل الخامس
المتغيرات العشوائية والتوزيعات الاحتمالية

Random Variables and Probability Distributions.

5 : 1 مقدمة

كثيراً ما تكون نتائج تجربة ما، أي نقاط فضاء العينة، قياسات وصفية أو نوعية، مثل الفضاء العيني لتجربة رمي قطعة نقود مرّة واحدة فإن النتيجة إما أن تكون صورة H أو كتابة T وهي قياسات نوعية ومثل ذلك رمي قطعة النقود مرتين فيكون فضاء العينة.

$$\Omega = \{HH, HT, TH, TT\}$$

وهي قياسات نوعية.

وهناك تجارب كثيرة تكون نقاط فضاء العينة فيها قياسات كمية أي قيماً عددية مثل رمي زهرة نرد مرّة واحدة وتسجيل العدد الظاهر إلى أعلى، ففي هذه الحالة يكون فضاء العينة

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

وفي جميع هذه الأنواع من التجارب، التي تكون النتائج فيها قياسات نوعية أو قياسات كمية فإن الأسئلة التي يمكن أن نطرحها تقتصر على نقاط فضاء العينة لأي تجربة. وبدلاً من الاكتفاء بطرح عدد محدود جداً من الأسئلة عن نقاط فضاء العينة، يمكننا أن نقرن قيماً عددية لتلك النقاط، أي نعرّف اقترانات حقيقية على فضاء العينة وبالتالي يكون لنا حرية كبيرة في طرح الأسئلة ويمكننا التعبير عن مفاهيم كثيرة نصوغها عن طريق الاقترانات.

5 : 2 المتغير العشوائي

إن الاقترانات الحقيقية التي نعرّفها على فضاء العينة لتجربة إحصائية تسمى متغيرات عشوائية وبالتالي فإن المتغير العشوائي يعيّن قيمة عددية لكل نتيجة بسيطة، (نقطة في الفضاء العيني).

تعريف (1)

المتغير العشوائي X هو اقتران حقيقي يعرّف على فضاء العينة.

سنرمز للمتغير العشوائي بحرف لاتيني كبير مثل X ولأي قيمة لذلك المتغير بحرف صغير x.

مثال (1) :

عرّف X بعدد ظهور الصورة H عند رمي قطعة نقود منتظمة 3 مرات.

من الواضح أن X متغير عشوائي حيث أنه يعيّن قيمة عددية لكل نقطة من نقاط فضاء العينة، ويمكننا كتابة ذلك كما في الجدول التالي:

النتيجة	قيمة X
HHH	3
HHT	2
HTH	2
HTT	1
THH	2

1	THT
1	TTH
0	TTT

لاحظ أن لكل نتيجة بسيطة (نقطة في فضاء العينة) يتعين قيمة واحدة فقط، ولكن من الممكن أن تأخذ عدة نتائج بسيطة القيمة العددية ذاتها، ولذلك يمكننا تجميع النتائج البسيطة التي تشترك في كل قيمة عددية من قيم المتغير العشوائي. ففي مثالنا نكتب:

الحادث المقابل	قيمة X
{TTT}	0
{HTT, THT, TTH}	1
{HHT, HTH, THH}	2
{HHH}	3

إن مجموعة القيم التي يأخذها المتغير العشوائي X تسمى فضاء X، والمجموعات الجزئية لهذا الفضاء تسمى حوادث ولكن معبّر عنها بدلالة X. ففي مثالنا، تكون الحوادث {X = 0}, {X = 1}, {X = 2}, {X = 3} وبدراسة المثال (1) يتضح أن الحوادث المرتبطة بالقيم المختلفة للمتغير X منفصلة عن بعضها البعض، وأن اتحادها يمثل فضاء العينة كلّه.

وكما ذكرنا سابقاً، فعلى فضاء العينة الواحد يمكننا تعريف عدد كبير من الاقترانات، أي عدد كبير من المتغيرات العشوائية.

مثال (2): عرّف المتغيرات العشوائية التالية على فضاء العينة في مثال (1):

أ. افرض Y تمثل الفرق المطلق بين عدد H وعدد T.

ب. افرض Z تمثل عدد H ناقصاً عدد T.

الحل: كما في المثال (1) نكتب قيم كل متغير والحوادث المقابلة لها.

أ)

الحادث المقابل	قيمة Y

{HHT, HTH, HTT, THH, THT, TTH}	1
{HHH, TTT}	3

ب)

الحادث المقابل	قيمة Z
{TTT}	-3
{HTT, THT, TTH}	-1
{HHT, HTH, THH}	1
{HHH}	3

يسمى المتغير العشوائي متغيراً منفصلاً Discrete إذا كانت القيم التي يأخذها إما محدودة أو لانهائية معدودة أي يمكن ربط قيمه واحداً إلى واحد مع مجموعة الأعداد الصحيحة الموجبة. إن الأمثلة السابقة كانت أمثلة للمتغير العشوائي المنفصل.

ومن جهة ثانية إذا كان المتغير العشوائي يمثل قياساً يأخذ جميع القيم في فترة فإنه يسمى متغيرا عشوائياً متصلاً (Continuous) ومن الأمثلة على ذلك درجة حرارة مريض، ارتفاع البنزين في خزان وقود سيارة، ارتفاع الماء في خزان وهكذا. ولا ننسى أن أدوات القياس التي لدينا لها درجة دقة معينة ولذلك عندما نحاول إيجاد قيمة متغير عشوائي متصل مثل طول الطالب في الصف العاشر، فإننا نستعمل أداة تقريبية للقياس مثل المسطرة أو المتر المدرّج لأقرب سم، وبالتالي فإن القياس المتصل (قياس المتغير العشوائي المتصل) يعتبر بأنه تجريد وأما القيمة التي نعطيها له فهي تقريبية.

وما دام المتغير العشوائي اقتراناً معرفاً على فضاء العينة فإن كل حادث في فضاء العينة يقابله مجموعة جزئية من فضاء المتغير العشوائي تسمى حادثاً بدلالة المتغير العشوائي وبالتالي يمكننا تعيين احتمالاً لهذا الحادث مساوياً لاحتمال الحادث في فضاء العينة.

مثال (3) :إذا كانت قطعة النقود في المثال (1) منتظمة ، أوجد :

P(X = 3), P(X = 2).

الحل : الحادث {X = 3} يقابل النتيجة HHH وبالتالي

$$P(X = 3) = P(HHH) = \left(\frac{1}{2}\right)^3 = \frac{1}{8}$$

الحادث {X = 2} يقابل الحادث:

A = {HHT, HTH, THH}

وبالتالي:

$$P(X = 2) = P(A) = 3 \times \left(\frac{1}{2}\right)^3 = \frac{3}{8}$$

ويمكننا وضع جدول يعطينا قيم X والاحتمالات المقابلة لها كما في الجدول (1):

قيمة X	الاحتمال P(X = x)
0	1/8
1	3/8
2	3/8
3	1/8

الجدول (1)

5 : 3 التوزيـــع الاحتمــالي المنفصــل Discrete Probability Distribution

بالرجوع إلى المثال (3) فقد كتبنا جدولا فيه قيم المتغير العشوائي واحتمال كل قيمة، ولبناء هذا الجدول عليك الرجوع إلى الفضاء العيني وعدِّ النقط التي تعطيك قيمة X المطلوبة، ومن ثم تحسب احتمال الحادث المكون من هذه النقط.

تعريف (2)

كل جدول أو معادلة يعطي جميع القيم التي يمكن أن يأخذها متغير عشوائي منفصل مع احتمال كل قيمة منها يسمى توزيعا احتماليا منفصلا، أو اقتراناً احتمالياً. لاحظ أن الجدول (1) مثال على التوزيعات الاحتمالية المنفصلة. ولاحظ أنك قد تضع الجدول على شكل معادلة وبالتالي يكون لديك:

أي معادلـة تحـدد احتمـال كـل قيمـة يأخـذها المتغيـر العشـوائي تسـمى توزيعا احتماليا منفصلا.

وأنت تلاحظ من التعريف أن التوزيع الاحتمالي المنفصل يحقق الشرطين التاليين.

(1) احتمال كل قيمة من قيم X غير سالب (2) مجموع الاحتمالات للقيم التي يأخذها X تساوي الواحد الصحيح.

وإذا عبرنا بالرمز f(x) للاحتمال P(X = x) فإن الشرطين السابقين يصبحان:

لجميع قيمx، $f(x) \geq 0$

جميع قيمx , $\sum f(x) = 1$

مثال (4) :
هل تمثل المعادلة:

$$f(x) = \frac{x}{15} \; ; x = 1, 2, 3, 4, 5$$

لغير ذلك 0 =;

توزيعا احتماليا؟

الحل: من الواضح أن $f(x) > 0$ للقيم $x = 1, 2, 3, 4, 5$ وهي صفر لغير هذه القيم، وبالتالي فإن الشرط (1) متحقق.

أما مجموع قيم f(x) فهي:

$$\sum_{x=1}^{5} f(x) = \sum_{x=1}^{4} \frac{x}{15} = \frac{1}{15} + \frac{2}{15} + \frac{3}{15} + \frac{4}{15} + \frac{5}{15} = 1$$

إذن فإن f(x) توزيع احتمالي، لأن الشرط (2) متحقق ويمكن وضع المعادلة السابقة في جدول (2) كما يلي:

جدول (2)

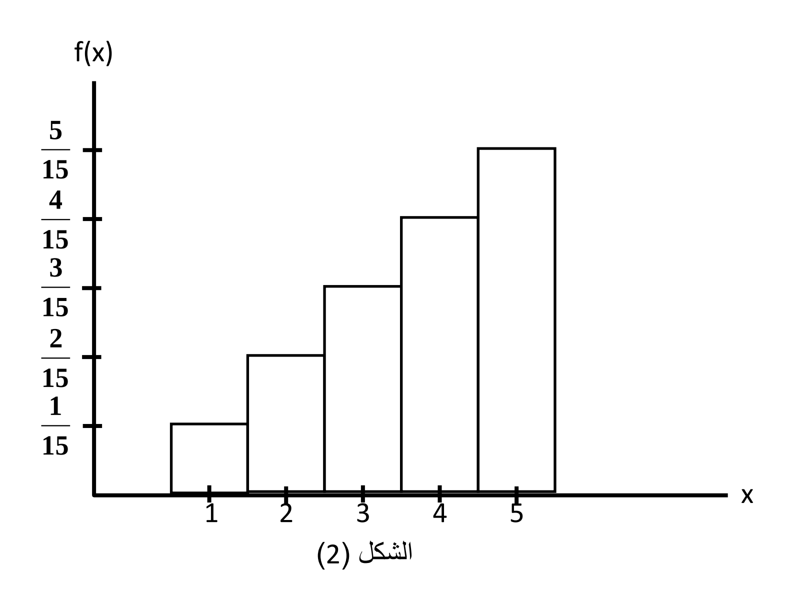
x	1	2	3	4	5
f(x)	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{3}{15}$	$\frac{4}{15}$	$\frac{5}{15}$

والآن، وقد تم تعريف التوزيع الاحتمالي على شكل جدول أو معادلة، فهل نتمكن من عرضه بيانيا؟ بالرجوع إلى عرض التوزيع التكراري بيانيا يمكنك عرض التوزيع الاحتمالي بواسطة المدرج الاحتمالي أو المضلع الاحتمالي (بنفس طريقة عرض التوزيع التكراري بواسطة المدرج التكراري أو المضلع التكراري) وذلك بوضع قيم X على المحور الأفقي ورسم مستطيلات تكون فيها كل قيمة من قيم X في منتصف قاعدة المستطيل ويكون ارتفاع المستطيل متناسبا مع احتمال تلك القيمة أي P(X = x).

مثال (5) :

أعرض الجدول (2) على شكل مدرج احتمالي.

الحل: ولعرض الجدول بيانيا نأخذ قيم X على المحور الأفقي وهي 1 ,2 ,3 ,4 ,5 ونرسم عليها مستطيلات على أن تكون كل قيمة في منتصف القاعدة ويكون ارتفاع كل مستطيل متناسبا مع احتمال تلك القيمة. كما يظهر في الشكل (2).



الشكل (2)

مثال (6) :

أوجد a في الجدول التالي التي تجعل ذلك الجدول يمثل توزيعا احتماليا واحسب احتمال X أكبر من 4

x	1	2	3	4	5	6
f(x)	a	1/5	1/5	1/20	3/20	1/10

الحل: $\sum_{x=1}^{6} f(x) = 1$

إذن $a + \frac{1}{5} + \frac{1}{5} + \frac{1}{20} + \frac{3}{20} + \frac{1}{10} = 1$

إذن $a + \frac{14}{20} = 1$

إذن $a = \frac{6}{20} = \frac{3}{10}$

احتمال X أكبر من 4 يساوي: $P(X = 5) + P(X = 6)$

$$= \frac{3}{20} + \frac{1}{10} = \frac{5}{20} = \frac{1}{4}$$

5 : 4 التوقع الرياضي Mathematical Expectation

(1) التوقع الرياضي ويسمى التوقع هو واحد من أهم المواضيع في الاحتمالات، ويعرّف بإضافته إلى المتغير أو اقتران المتغير فنقول التوقع الرياضي للمتغير العشوائي أو التوقع الرياضي للاقتران g(X). أما التوقع الرياضي للمتغير العشوائي X أو توقع X أو معدل X فهو قيمة عددية تصف مركز ذلك المتغير.

ونكتفي في هذا الفصل بدراسة التوقع الرياضي في حالة المتغير العشوائي المنفصل.

تعريف (3)

التوقع الرياضي للمتغير العشوائي المنفصل الذي اقترانه الاحتمالي f(x) هو المقدار.

$$\mu = E(X) = \sum_{x} x f(x)$$

(شريطة أن $\sum |x|\, f(x) < \infty$ أي أن المتسلسلة تتقارب بشكل مطلق).

مثال (7)

أوجد التوقع للمتغير X الذي توزيعه الاحتمالي كما في الجدول التالي :

x	f(x)	x . f(x)
3	0.3	0.9
4	0.2	0.8
5	0.2	1.0
6	0.1	0.6
7	0.2	1.4
		4.7

الحل: اضرب القيمة x في f(x) وضع الناتج في العمود الثالث في الجدول (4) اجمع الأعداد في ذلك العمود فتجد $\mu = 4.7$

مثال (8) :

أوجد توقع X إذا كان

$f(x) = \frac{x}{15}$; x = 1, 2, 3, 4, 5

لغير ذلك ; 0 =

الحل: $\mu = E(X) = \sum_x x.f(x)$

$$1\times\frac{1}{15}+2\times\frac{2}{15}+3\times\frac{3}{15}+4\times\frac{4}{15}+\frac{5}{15}\times 5$$

$$=\frac{1+4+9+16+25}{15}=\frac{55}{15}=\frac{11}{3}$$

(2) خواص التوقع الرياضي Properties of Mathematical Expectation

نظرية (1)

لكل متغير عشوائي X، إذا كان b, a ثابتين فإن:

$E(a X + b) = a E(X) + b$

وهذا يعني أنه إذا ضربت المتغير العشوائي X في عدد ثابت، وأضفت إليه عددا ثابتا، فإن التوقع يتأثر بنفس الطريقة فيصبح التوقع بعد التعديل مساويا للتوقع الأصلي مضروبا في العدد a ومضافا إليه العدد b.

مثال (9) :

إذا كان E(X) = 6، أوجد E(3 X + 5).

الحل:

$$E(3X + 5) = 3E(X) + 5$$

$$3 \times 6 + 5 = 23$$

مثال (10) :

جد توقع b إذا كان b عددا ثابتا.

الحل: ضع a = 0 في نظرية (1) تجد

$$E(0 \times X + b) = 0.E(X) + b = b$$

إذاً $E(b) = b$ إذا كان b عدداً ثابتاً، أي أن توقع العدد الثابت يساوي نفس العدد الثابت

كثيرا ما نحتاج لإيجاد توقع اقتران متغير عشوائي، كأن نقول ما هو توقع X^2 أو ما هو توقع X^3 أو بشكل عام ما هو توقع H(X) حيث H هو اقتران في X.

تعريف (4) :

إذا كان X متغيرا عشوائيا منفصلا وكان توزيعه الاحتمالي f(x) وكان H(X) اقترانا في X فإن توقع H(X) هو

$$E[H(X)] = \sum_x H(x) \, . \, f(x)$$

شريطة أن يكون $\sum_x H(x) \, f(x) < \infty$.

مثال (11) :

أوجد توقع X^2 إذا كان

$$f(x) = \frac{x}{15} \; ; x = 1, 2, 3, 4, 5$$

$= 0$; لغير ذلك

الحل: طبق التعريف الأخير وذلك بوضع x^2 بدلا من H(x) تحصل على:

$$E(X^2) = \sum_x (x^2) \, . \, f(x)$$

$$= 1^2 \times \frac{1}{15} + 2^2 \times \frac{2}{15} + 3^2 \times \frac{3}{15} + 4^2 \times \frac{4}{15} + 5^2 \times \frac{5}{15}$$

$$= \frac{1}{15} \; (1 + 8 + 27 + 64 + 125) = \frac{225}{15} = 15$$

إذا كان μ توقع متغير عشوائي X فإن توقع الاقتران $(X - \mu)^2$ له أهمية خاصة في دراسة الإحصاء، ولذلك إليك هذا التعريف:

تعريف (5)

إذا كان μ توقع متغير عشوائي X فإن توقع $(X - \mu)^2$ يسمى تباين X ويعبر عنه بالرمز σ^2، أي أن:

تباين المتغير العشوائي X هو $\sigma^2 = E(X - \mu)^2$

وإذا كان f(x) التوزيع الاحتمالي للمتغير X

أصبح لديك: تباين X هو

$$\sigma^2 = \sum_x (X-\mu)^2 f(x)$$

شريطة أن $\sum_x (X-\mu)^2 f(X) < \infty$

لاحظ أن تباين X يمكن تعريفه من تعريف (4) بوضع

$$H(X) = (X - \mu)^2$$

هناك حالات خاصة أيضاً منها

$$H(X) = X^k$$

وفي هذه الحالة نحصل على $E[H(X)] = E(X^k)$ وهو العزم k حول الصفر فمثلا $E(X^2)$ هو العزم الثاني للمتغير X.

مثال (12) :

جد تباين X إذا كان توزيعه الاحتمالي كما يلي:

x	f(x)
10	$\frac{1}{4}$

20	$\frac{1}{4}$
30	$\frac{1}{4}$
40	$\frac{1}{4}$

الحل: أولا، جد توقع X أي (معدل X):

$$\mu = E(X) = 10 \times \frac{1}{4} + 20 \times \frac{1}{4} + 30 \times \frac{1}{4} + 40 \times \frac{1}{4} = 25$$

والآن، طبّق التعريف

$$\sigma^2 = \sum (x-\mu)^2 .f(x) = (10-25)^2 \times \frac{1}{4} + (20-25)^2 \times \frac{1}{4} + (30-25)^2$$

$$\times \frac{1}{4} + (40-25)^2 \times \frac{1}{4} = \frac{225+25+25+225}{4} = 125$$

أما الجذر التربيعي الموجب للتباين فيسمى الانحراف المعياري أي أن الانحراف المعياري للمتغير X هو $\sigma = \sqrt{\sigma^2}$.

ويمكننا إيجاد توقع X وتباين X بإيجاد المجاميع المطلوبة في جدول واحد كما يتضح في المثال التالي:

مثال (13) : رميت زهرتا نرد منتظمتان مرة واحدة. افرض X يمثل مجموع العددين الظاهرين إلى أعلى. اكتب فضاء العينة وأوجد التوزيع الاحتمالي للمتغير X ثم أوجد معدل X أي E(X) وتباين X.

الحل: يوجد 36 نقطة في فضاء العينة كما يظهر في الجدول التالي:

الوجه الأول / الوجه الثاني	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)

2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

نرتب الحل كما في الجدول (3) حيث يمثل العمودان (1)، (2) قيم المتغير العشوائي والاحتمالات المقابلة لها، أي التوزيع الاحتمالي للمتغير X، ومجموع العمود (3) يعطي معدل X وهو $\mu = 7$

ومجموع عمود (6) يعطي التباين σ^2 وهو $\sigma^2 = \frac{35}{6}$

قيمة x	f(x)	xf(x)	$(x-\mu)$	$(x-\mu)^2$	$(x-\mu)^2 f(x)$
2	1/36	2/36	-5	25	25/36
3	2/36	6/36	-4	16	32/36
4	3/36	12/36	-3	9	27/36
5	4/36	20/36	-2	4	16/36
6	5/36	30/36	-1	1	5/36
7	6/36	42/36	0	0	0
8	5/36	40/36	1	1	5/36
9	4/36	36/36	2	4	16/36
10	3/36	30/36	3	9	27/36
11	2/36	23/36	4	16	32/36
12	1/36	12/36	5	25	25/36
المجموع	1	7			35/6

الجدول (3)

ما هو تأثير التحويل الخطي على التباين؟ نجيب عن هذا السؤال بالنظرية التالية:

نظرية (2) :

إذا كان X متغيراً عشوائياً منفصلاً معدله μ وتباينه σ^2_x وكان $Y = aX + b$ حيث a، b ثابتان فإن

(1) $\mu_Y = E(Y) = a\,E(X) + b$

(2) $\sigma^2_Y = Var.\,Y = a^2\sigma^2_x$

(3) الانحراف المعياري للمتغير Y هو

$\sigma_Y = |a|\,.\,\sigma_x$

مثال (14) :

إذا كان للمتغير X، $\mu_x = 50$، $\sigma^2_x = 16$ أوجد معدل Y وتباينه وانحرافه المعياري إذا كان $Y = 3X - 4$

الحل:

$$\mu_Y = E(3X - 4) = 3\mu_x - 4 = 3 \times 50 - 4 = 146$$

$$\sigma^2_Y = 3^2\sigma^2_x = 9 \times 16 = 144$$

$$\sigma_Y = \sqrt{144} = 12$$

يمكن حساب التباين باستعمال العزم الثاني بدلاً من استعمال الانحرافات $(X - \mu)$ وذلك حسب النظرية:

نظرية (3) : إذا كان X متغيرا عشوائيا معدله μ وتباينه σ^2 فإن :

$$\sigma^2 = E(X^2) - \mu^2$$

البرهان : افرض أن التوزيع الاحتمالي ل X هو f(x) ، وأن التوزيع منفصل . من التعريف :

$$\begin{aligned}
\sigma^2 &= E\left((X-\mu)^2\right) \\
&= \sum_{\text{x لجميع}} (x-\mu)^2 f(x) \\
&= \sum_{\text{x لجميع}} (x^2 - 2\mu + \mu^2) f(x) \\
&= \sum_{\text{x لجميع}} x^2 f(x) - 2\mu \sum_{\text{x لجميع}} x f(x) + \mu^2 \sum_{\text{x لجميع}} f(x) \\
&= E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2
\end{aligned}$$

حيث $\sum_{\text{x لجميع}} f(x) = 1 \quad , \quad \sum_{\text{x لجميع}} x f(x) = \mu$

5 : 5 توزيعات احتمالية خاصة

Special Probability Distributions

هناك كثير من المتغيرات العشوائية المنفصلة التي شاع استعمال توزيعاتها الاحتمالية بحيث صار من المفيد دراسة كل منها على حده، ومن هذه التوزيعات

(1) التوزيع المنفصل المتجانس Uniform Discrete Distribution

إذا أخذ متغير عشوائي عدداً محدوداً من القيم وكان احتمال أي قيمة مساوياً لاحتمال أي قيمة أخرى نقول إن توزيع هذا المتغير التوزيع المتجانس المنفصل. ومن التجارب البسيطة التي تؤدي إلى التوزيع المتجانس المنفصل تجربة سحب ورقة من مجموعة أوراق عددها n ومكتوب عليها الأعداد $x_1, x_2, ..., x_n$ فإذا فرضنا أن X يمثل العدد المكتوب على الورقة المسحوبة فإن توزيع X هو التوزيع المتجانس ويكون

$$f(x_i) = P(X = x_i) = \frac{1}{n}\ ; i = 1, 2, ... , n$$

تعريف (6) : لأي عدد صحيح موجب n فإن الاقتران:

$$f(x_i) = \frac{1}{n}, i = 1, 2, ... , n$$

يسمى توزيعاً احتمالياً متجانساً منفصلاً.

ويسمى المتغير X الذي توزيعه التوزيع المتجانس المنفصل متغيراً متجانساً منفصلاً.

مثال (15): رميت زهرة نرد منتظمة مرّة واحدة.

افرض X يمثل العدد الظاهر إلى أعلى. ما هو التوزيع الاحتمالي للمتغير X.

الحل: يوجد في فضاء العينة 6 نقاط أي:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

وبما أن زهرة النرد منتظمة فإن احتمال ظهور أي عدد يساوي احتمال ظهور أي عدد آخر

إذاً $\frac{1}{6}$ $f(i) = P(X = i) =\ ; i = 1, 2, 3, 4, 5, 6$

أي أن توزيع X توزيع منفصل متجانس.

نظرية (4)

إذا كان توزيع X توزيعاً منفصلاً متجانساً معطى بالاقتران

$$f(x) = \frac{1}{n}, \; x = 1, 2, \ldots, n$$

فإن معدل X أي توقع X يكون

$$\mu = E[X] = \frac{n+1}{2}$$

وتباين X يكون $\sigma^2 = \frac{n^2-1}{12}$

مثال (16) :

أوجد معدل X وتباينه في مثال (15)

الحل:

$$\mu = E[X] = \frac{6+1}{2} = \frac{7}{2}$$

$$\sigma^2 = \frac{n^2-1}{12} = \frac{6^2-1}{12} = \frac{35}{12}$$

لاحظ أن بإمكانك إيجاد μ, σ^2 مباشرة من التعريف

$$\mu = E[X] = \sum_{i=1}^{6} i \times \frac{1}{6} = \frac{7}{2}$$

$$\sigma^2 = E[X^2] - \mu^2$$

$$E(X^2) = \sum_{i=1}^{6} i^2 \times \frac{1}{6} = \frac{1}{6}[1 + 4 + 9 + 16 + 25 + 36] = \frac{91}{6}$$

إذاً

$$\sigma^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}$$

(2) توزيع ذات الحدين

Binomial Distribution

في كثير من التجارب تكون نتيجة كل محاولة للتجربة أحد الأمرين، إما نجاح وإما فشل. وتتألف هذه التجارب من تكرار وإعادة المحاولات المستقلة عن بعضها البعض، فمثلا عند رمي قطعة نقود فإن النتيجة تكون إما ظهور صورة أو ظهور كتابة، وتكون نتيجة أي محاولة مستقلة عن نتيجة أي محاولة أخرى فإذا كانت نتيجة المحاولة الأولى ظهور الصورة إلى أعلى فإن ذلك لا يؤثر في نتيجة محاولة الثانية وهكذا. إن هذه المحاولات تسمى محاولات بيرنوللي Bernoulli Trials.

تعريف (7) محاولات بيرنوللي

كل تجربة تحقق الشروط التالية تسمى محاولة بيرنوللي.

1. نتيجة كل محاولة أحد ناتجين، نسمي أحدهما "نجاحا" والآخر "فشلا".

2. نتيجة كل محاولة مستقلة عن نتيجة أي محاولة أخرى.

3. احتمال النجاح في كل محاولة ثابت وليكن p.

ولذلك فاحتمال الفشل ثابت وهو $q = 1 - p$

إن المحاولات التي تعطى أحد ناتجين فقط، إما "نجاح" وإما "فشل" تحدث في حالات أخذ العينات من المجتمعات التي تكون عناصرها مصنّفة إلى صنفين، نجاح وفشل. ومن الأمثلة على ذلك :

(أ) إذا كان لديك مجموعة من المصابيح الكهربائية وأردت فحصها واحداً تلو الآخر. إن كل محاولة هي محاولة بيرنوللي لأنها تحقق الشروط الثلاثة المذكورة آنفاً.

(ب) فحص سجلات الولادة في أحد المستشفيات لمعرفة نسبة الولادات بالعملية القيصرية.

(جـ) فحص مجموعة من طلاب الصف العاشر لمعرفة نسبة من يحتاج إلى استعمال نظارة طبيّة.

إذا فرضنا أن X هو نتيجة محاولة بيرنوللي التي احتمال النجاح فيها p وإذا فرضنا $X = 1$ إذا كانت النتيجة "نجاحاً"، $X = 0$ إذا كانت النتيجة "فشلاً" فإننا نحصل على التوزيع الاحتمالي للمتغير X كما في الجدول (4)

x	f(x) = P(X = x)
1	p
0	1 - p

الجدول (4)

والآن، إذا كررت محاولة بيرنوللي عدداً ثابتاً من المرات وليكن n فإننا نحصل على تجربة ذات الحدين Binomial Experiment أي أن تجربة ذات الحدين هي كل تجربة تحقق الشروط الثلاثة لتجربة بيرنوللي بالإضافة إلى الشرط الرابع وهو أن نجري محاولات بيرنوللي عددا ثابتا من المّرات n.

تعريف (8)

إذا أجريت تجربة بيرنوللي n من المرات وكان احتمال "النجاح" في المحاولة الواحدة p وكان X يمثل عدد "النجاح" في المحاولات كلها فإن

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x} \quad ; \qquad x = 0, 1, 2, \ldots, n$$

وهذا هو التوزيع الاحتمالي لمتغير ذات الحدين ونمثله بالرمز b(x; n, p) وبذلك يكون لدينا:

$$b(x; n, p) = P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x} \; ; \qquad x = 0, 1, 2, \ldots, n$$

مثال (17) :

رميت قطعة نقد متزنة أربع مرات.

جد التوزيع الاحتمالي لعدد ظهور الصورة في هذه التجربة.

الحل: إن هذه التجربة تحقق شروط تجربة ذات الحدين حيث $p = \frac{1}{2}$, $n = 4$, وبوضع X = عدد ظهور الصورة في المحاولات الأربع، ومن التعريف الأخير تجد:

$$b(x; 4, \frac{1}{2}) = P(X = x)$$

$$= \binom{4}{x}\left(\frac{1}{2}\right)^{x}\left(1-\frac{1}{2}\right)^{4-x} ; x = 0, 1, 2, 3, 4.$$

$$= \binom{4}{x}\left(\frac{1}{2}\right)^{4} ; x = 0, 1, 2, 3, 4$$

فلو حسبت هذه القيم $b(x; 4, \frac{1}{2})$ لقيم x المختلفة لوجدت:

$$b(0, 4, \frac{1}{2}) = \binom{4}{0}\left(\frac{1}{2}\right)^{4} = \frac{1}{16}$$

$$b(1, 4, \frac{1}{2}) = \binom{4}{1}\left(\frac{1}{2}\right)^{4} = \frac{4}{16}$$

$$b(2, 4, \frac{1}{2}) = \binom{4}{2}\left(\frac{1}{2}\right)^{4} = \frac{6}{16}$$

$$b(3, 4, \frac{1}{2}) = \binom{4}{3}\left(\frac{1}{2}\right)^{4} = \frac{4}{16}$$

$$b(4,4, \frac{1}{2}) = \binom{4}{4}\left(\frac{1}{2}\right)^{4} = \frac{1}{16}$$

من الواضح أن كل قيمة هي كسر ما بين الصفر والواحد ومجموعها يساوي 1 أي أن

$P(X = x) \geq 0$

$$\sum_{x} P(X = x) = 1$$

وهذا يعني أن $b(x; 4, \frac{1}{2})$ هو اقتران احتمالي لأنه يحقق شروط هذا الاقتران.

مثال (18) :

رميت زهرة نرد منتظمة 3 مرات، ما احتمال عدم ظهور العدد 1 فيها؟ ما احتمال ظهور العدد 1 مرتين؟

الحل: عند رمي زهرة النرد مرة واحدة

فإن الفضاء العيني $\Omega = \{1, 2, 3, 4, 5, 6\}$

واحتمال ظهور 1 يساوي $\frac{1}{6}$

واحتمال عدم ظهور 1 يساوي $\frac{5}{6}$

إن هذه التجربة تحقق شروط تجربة ذات الحدين حيث

$$q = 1 - p = \frac{5}{6}, \; p = \frac{1}{6}, \; n = 3$$

ضع X = عدد ظهور العدد 1 في المحاولات الثلاثة وبالتالي فإن التوزيع الاحتمالي للمتغير X يكون:

$$b(x; 3, \frac{1}{6}) = \binom{3}{x}\left(\frac{1}{6}\right)^{x}\left(\frac{5}{6}\right)^{3-x} \quad x = 0, 1, 2, 3$$

إن عدم ظهور 1 يعني أن $X = 0$ ولذلك

$$P(X = 0) = b(0; 3, \frac{1}{6})$$

$$= \binom{3}{0}\left(\frac{1}{6}\right)^{0}\left(\frac{5}{6}\right)^{3-0}$$

$$= \left(\frac{5}{6}\right)^{3} = 0.579$$

وظهور 1 مرتين يعني أن X = 2 ولذلك

$$P(X = 2) = b(2; 3, \frac{1}{6}) = \binom{3}{2}\left(\frac{1}{6}\right)^{2}\left(\frac{5}{6}\right)^{3-2}$$

$$= 3.\left(\frac{1}{6}\right)^{2}.\left(\frac{5}{6}\right) = \frac{15}{216} = 0.069$$

مثال (19) :

إذا كانت نسبة التالف من المصابيح الكهربائية في مصنع تساوي 0.001، وأخذت عينة حجمها 10 مصابيح بطريقة عشوائية، ما احتمال أن يكون عدد التالف في هذه العينة يساوي صفراً ؟ ما احتمال أن يكون عدد التالف اثنين؟

الحل: من الواضح أن المصباح إما أن يكون صالحاً أو تالفاً. إذا وضعت p = 0.001 وهو احتمال أن يكون المصباح تالفا، نلاحظ أن أخذ 10 مصابيح من المصنع يعني إجراء تجربة ذات الحدين فيها n = 10 :p = 0.001

ضع X = عدد المصابيح التالفة ويكون التوزيع الاحتمالي b(x; 10, 0.001) ويكون احتمال عدد التالف صفرا هو:

$$\mathbf{P(X = 0) = b(0; 10, 0.001)}$$

$$= \binom{10}{0}(0.001)^{0}(0.999)^{10-0} = (0.999)^{10}$$

أما احتمال عدد التالف اثنين فهو:

$$\mathbf{P(X = 2) = b(2; 10, 0.001)}$$

$$= \binom{10}{2} (0.001)^2 (0.999)^{10-2} = 45\ (0.999)^8 \cdot 10^{-6}$$

تحتاج في كثير من الأحيان إلى حساب احتمال حوادث مثل $X \geq 1$ أو $X \leq 7$ أو غير ذلك في مثل هذه الحالات يمكنك استعمال قوانين الاحتمال ومفهوم جمع الاحتمالات كما في الأمثلة التالية:

مثال (20) :

في تجربة ذات الحدين $n = 15$، $p = 0.3$

أوجد احتمال أن X أقل من أو يساوي 2 حيث X هو عدد "النجاح".

الحل: المطلوب إيجاد $P(X \leq 2)$

إن هذا يعني أن $X = 0$ أو $X = 1$ أو $X = 2$ وبالتالي

$$\mathbf{P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)}$$

$$\mathbf{P(X \leq 2) = b(0; 15, 0.3) + b(1; 15, 0.3) + b(2; 15, 0.3)}$$

$$= \binom{15}{0} (0.3)^0 (0.7)^{15} + \binom{15}{1} (0.3)^1 (0.7)^{14} + \binom{15}{2} (0.3)^2 (0.7)^{13}$$

$$= (0.7)^{15} + 15\ (0.3)\ (0.7)^{14} + 105\ (0.3)^2\ (0.7)^{13}$$

$$= 0.0047 + 0.0305 + 0.3052$$

$$= 0.3404$$

مثال (21): X يخضع لتوزيع ذات الحدين فيه $n = 10$، $p = 0.1$ أوجد $P(X \leq 9)$.

الحل: لو أردنا إيجاد هذا الاحتمال مباشرة وجب أن نجد مجموع عشرة حدود هي

$$P(X = 0) + P(X = 1) + P(X = 2) + ... + P(X = 9)$$

وهذا يحتاج لجهد كبير ولكن نعلم أن

$$P(X \leq 9) = 1 - P(X = 10)$$

لأن الحادث $\{X = 10\}$ متمم للحادث $\{X \leq 9\}$

$$P(X \le 9) = 1 - \binom{10}{10}(0.1)^{10}(0.9)^{0} = 1 - (0.1)^{0}$$

أما إذا كان $\mathbf{P = \frac{1}{3}, n = 10}$ فإن :

$$\mathbf{P(X \le 9) = 1 - P(X = 10)}$$

$$= 1 - \binom{10}{10}\left(\frac{1}{3}\right)^{10} \cdot \left(\frac{2}{3}\right)^{10-10}$$

$$= 1 - \left(\frac{1}{3}\right)^{10} = 0.99998 \approx 1$$

بعد إيجاد التوزيع الاحتمالي للمتغير العشوائي المنفصل يمكننـا تمثيل ذلك التوزيع الاحتمالي بيانيا بالمدرج الاحتمالي أو بواسطة الأعمدة.

مثال (22) :

إذا كان X متغيرا عشوائيا لذات الحدين حيث $n = 5, p = \frac{1}{2}$

ضع قيم X والاحتمالات المقابلة لها في جدول ثم اعرض ذلك بيانيا بواسطة المدرج الاحتمالي وبواسطة الأعمدة.

الحل:

$$P(X = x) = b(x, 5, \frac{1}{2})$$

حيث x تأخذ القيم 0, 1, 2, 3, 4, 5

$$P(X = 0) = b(0; 5, \frac{1}{2}) = \binom{5}{0}\left(\frac{1}{2}\right)^{0}\left(\frac{1}{2}\right)^{5-0} = \left(\frac{1}{2}\right)^{5} = \frac{1}{32} = 0.031$$

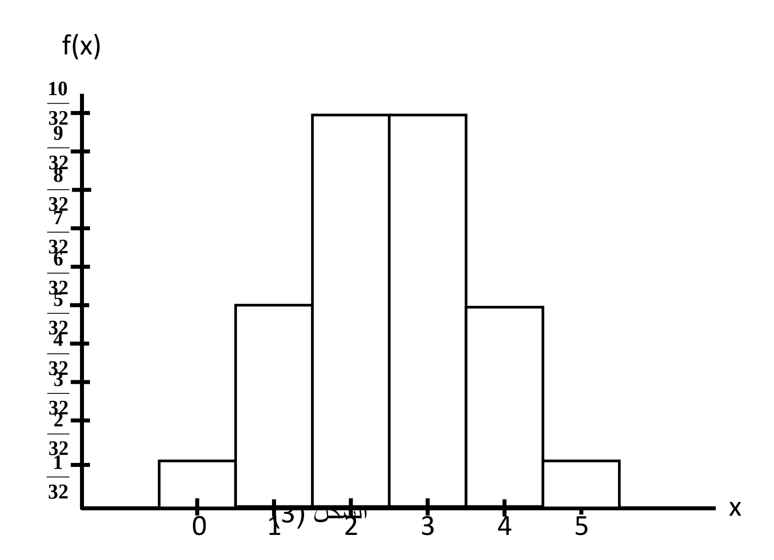
$$P(X = 1) = b(1; 5, \frac{1}{2}) = \binom{5}{1}\left(\frac{1}{2}\right)^{1}\left(\frac{1}{2}\right)^{5-1}$$

$$= 5\left(\frac{1}{2}\right)^5 = \frac{5}{32} = 0.156$$

وهكذا نحصل على الجدول (5) :

قيمة x	0	1	2	3	4	5
P(X = x)	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{10}{32}$	$\frac{10}{32}$	$\frac{5}{32}$	$\frac{1}{32}$

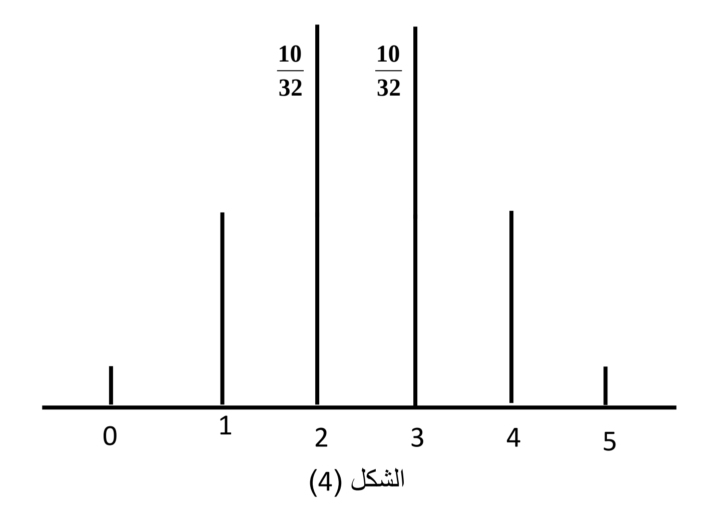
المدرج الاحتمالي يظهر في الشكل (3) والعرض بواسطة الأعمدة يظهر في الشكل (4).



الشكل (3)

$\frac{5}{32}$ $\frac{5}{32}$

$\frac{1}{32}$ $\frac{1}{32}$



الشكل (4)

باستعمال التعاريف الخاصة بالتوقع والتباين للمتغير العشوائي، وتعريف التوزيع الاحتمالي لذات الحدين، يمكن برهنة النظرية التالية:

نظرية (4)

إذا كان X متغيرا عشوائيا لذات الحدين توزيعه الاحتمالي $b(x; n, p)$ فإن معدل X أي (توقع X) يساوي $\mu = E[X] = np$

وتباين X يساوي $\sigma^2 = np(1 - p)$

مثال (23) :

تُفحص أجهزة الحاسوب قبل اعتبارها صالحة وتسليمها إلى المشترين. فإذا كان نسبة نجاح هذه الأجهزة في الفحص 0.95 وأرسل 30 جهازاً للفحص، فكم جهازاً تتوقع أن تكون صالحة ، وما هو الانحراف المعياري لها؟

الحل: افرض عدد الأجهزة الناجحة X

فيكون التوزيع الاحتمالي للمتغير X هو

$b(x; 30, 0.95)$

من النظرية السابقة $\mu_x = 30 \times 0.95 = 28.5$

وهو عدد الأجهزة التي يتوقع أن تنجح

التباين يساوي $\sigma^2_x = 30 \times 0.95 \times 0.05 = 1.425$

والانحراف المعياري يساوي

$$\sigma_x = \sqrt{1.425} = 1.19$$

مثال (24) :

ما هو التوقع والتباين لمتغير ذات الحدين إذا كان $p = \frac{2}{3}$, n = 60

الحل:

$$\mu_x = np = 60 \times \frac{2}{3} = 40$$

$$\sigma^2_x = np\,(1 - p) = 60 \times \frac{2}{3} \times \frac{1}{3} = 13\frac{1}{3}$$

بالعودة إلى خواص تجربة ذات الحدين نرى أن توزيع ذات الحدين عادة ما يرتبط بمفهوم الاختيار أو السحب مع الإرجاع. إن مفهوم السحب العشوائي مع الإرجاع يتلخص في أنه إذا أردت اختيار n عنصراً من مجموعة عناصر أو مجتمع بحيث يكون احتمال اختيار أي عنصر مساوياً لاحتمال اختيار أي عنصر آخر وبحيث كلما سحبنا عنصراً وشاهدناه نرجعه ثانية إلى المجموعة الأصلية أي أن صفات المجتمع الأصلي من حيث عدد العناصر وأنواعها تبقى ثابتة.

مثال (25) :

صندوق فيه 6 كرات حمراء و 4 كرات بيضاء، سحبت منه 5 كرات مع الإرجاع. افرض X يمثل عدد الكرات الحمراء. ما هو التوزيع الاحتمالي للمتغير X.

الحل: بما أن السحب مع الإرجاع فذلك يعني أنه كلما أخذنا كرة نلاحظ لونها ثم نرجعها إلى الصندوق وبالتالي يبقى دائماً في الصندوق 6 كرات حمراء، 4 كرات بيضاء. فإذا عبّرنا عن نتيجة سحب الكرة الحمراء بصفة "نجاح" يكون سحب الكرة البيضاء صفة "فشل".

احتمال الحصول على كرة حمراء في أي محاولة هو $\frac{6}{10}$ وكل محاولة مستقلة عن أي محاولة أخرى، إذاً فالمحاولات هي محاولات بيرنوللي وعندما نكررها 5 مرات نحصل على تجربة ذات الحدين ويكون توزيع X هو $b(x; 5, 0.6)$ حيث $x = 0, 1, 2, 3, 4, 5$.

(3) التوزيع فوق الهندسي The Hypergeometric Distribution

بينما يرتبط توزيع ذات الحدين بالسحب مع الإرجاع فإن توزيع فوق الهندسي يرتبط بالسحب دون الإرجاع والذي تعريفه: نقول إننا استعملنا السحب العشوائي دون الإرجاع لاختيار عينة حجمها n من مجموعة عناصر أو مجتمع إذا اخترنا مجموعة جزئية حجمها n بطريقة عشوائية من مجموعة العناصر بحيث أن أي مجموعة حجمها n يكون لها نفس الاحتمال لتكون مختارة مثل أي مجموعة أخرى ذات الحجم نفسه.

افرض أن لدينا صندوقا فيه 6 كرات حمراء و 4 كرات بيضاء وكلها متشابهة. أردنا سحب 3 كرات منها على التوالي دون إرجاع والمطلوب معرفة احتمال أن تكون إحدى هذه الكرات حمراء.

من المعلوم أن نسبة الكرات الحمراء عند بدء السحب تساوي $\frac{6}{10}$ أي 0.6 ولذلك فإن احتمال أن تكون أول كرة قمت بسحبها حمراء = $\frac{6}{10}$ وبعد سحب هذه الكرة فإن احتمال أن تكون الكرة الثانية حمراء يعتمد على لون الكرة الأولى، ويكون هذا الاحتمال احتمالا مشروطا، فإذا كانت الأولى حمراء فاحتمال أن تكون الثانية حمراء يساوي $\frac{5}{9}$، أما احتمال أن تكون الثانية حمراء إذا علم أن الأولى بيضاء فهو $\frac{6}{9}$.

إن هذا يعني أن احتمال سحب كرة حمراء في كل محاولة غير ثابت لأنه في كل مرّة يتغير عدد الكرات، وبالتالي فإن تجربة ذات الحدين غير متحققة والسبب في ذلك أن طريقة سحب الكرات التي قمت بها كانت طريقة السحب بدون إرجاع، أي أنك عندما تسحب الكرة الأولى فإنك تشاهد لونها وتطرحها جانبا ثم تسحب الكرة الثانية وهكذا.

إن السحب بدون إرجاع مرتبط بالتجربة فوق الهندسية.

إن التجربة فوق الهندسية هي تلك التجربة التي تحقق الشرطين التاليين:

1. يتم اختيار عينة عشوائية حجمها n من مجتمع فيه N من العناصر ويكون سحب العينة بدون إرجاع.

2. إن المجتمع المذكور فيه N_1 من العناصر من نوع معين نسميها "نجاحا" وباقي العناصر N_2 من نوع آخر نسميها "فشلا"، أي أن $N = N_1 + N_2$. إذا فرضت أن عدد النجاحات التي تحصل عليها في التجربة فوق الهندسية هو X فإن X يسمى متغيرا عشوائيا فوق هندسي.

أما التوزيع الاحتمالي للمتغير العشوائي فوق الهندسي فيسمى التوزيع فوق الهندسي، واحتمال الحصول على x من النجاحات في هذه الحال يعبر عنه بالرمز $h(x; n, N_1, N_2)$ حيث يعتمد هذا الاحتمال على حجم المجتمع N وحجم العينة n وعدد النجاحات في المجتمع N_1 وعدد الفشل N_2.

مثال (26) :صف فيه 10 طالبات و 15 طالباً، أردنا تشكيل لجنة من 3 من هذا الصف. أوجد التوزيع الاحتمالي لعدد الطالبات في اللجنة.

الحل: افرض X يمثل عدد الطالبات في اللجنة. من الواضح أن هذه التجربة تحقق الشرطين للتجربة فوق هندسية، حيث $N_1 = 10$, $N_2 = 15$, $N = 25$ ، $n = 3$

ومن الواضح أن عدد الطالبات في اللجنة يمكن أن يكون 0 أو 1 أو 2 أو 3 باستعمال قاعدة التوافيق، يمكن اختيار 3 من 25 بطرق عددها $\binom{25}{3}$.

ويمكن عدم اختيار أي طالبة من بين 10 طالبات بطرق عددها $\binom{10}{0}$ ويمكن

اختيار 3 طلاب. من بين 15 طالباً بطرق عددها $\binom{15}{3}$.

$$P(X = 0) = h(0; 3, 10, 15)$$

$$= \frac{\binom{10}{0}\binom{15}{3}}{\binom{25}{3}} = \frac{455}{2300} = 0.198$$

وبنفس المنطق تجد احتمال اختيار طالبة وطالبين هو

$$P(X = 1) = h(1; 3, 10, 15)$$

$$= \frac{\binom{10}{1}\binom{15}{2}}{\binom{25}{3}} = \frac{1050}{2300} = 0.457$$

وبنفس الطريقة نحسب احتمال اختيار طالبتين وطالب واحتمال اختيار ثلاث طالبات وصفر طالباً.

ويمكن وضع التوزيع فوق الهندسي للمتغير X في جدول كالتالي:

3	2	1	0	x
0.052	0.293	0.457	0.198	P(X = x)

كما يمكن إعطاء هذا التوزيع على شكل معادلة:

$$P(X = x) = \frac{\binom{10}{x}\binom{15}{3-x}}{\binom{25}{3}} \; ; \; x = 0, 1, 2, 3$$

تعريف (8) :

إذا كان مجتمع حجمه N يحتوي على N_1 عنصرا نسميها "نجاحا" $N_2 = (N - N_1)$ عنصرا نسميها "فشلا"، واخترنا عينة حجمها n دون إرجاع من هذا المجتمع وفرضنا أن عدد النجاحات في هذه العينة التي حجمها n هو المتغير X، فإن التوزيع الاحتمالي للمتغير X هو التوزيع فوق الهندسي، ويعطى بالمعادلة:

$$h(x; n, N_1, N_2) = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}} \quad , \quad x = 0, 1, \ldots, n$$

إذا كانت $N_1 \geq n$ ولكن إذا كانت $N_1 < n$ فإن x تأخذ القيم $0, 1, 2, \ldots, N_1$

مثال (27) :

صندوق فيه 13 كرة حمراء، 7 كرات بيضاء، أخذت منه 4 كرات دون إرجاع، جد التوزيع الاحتمالي لعدد الكرات البيضاء في العينة المسحوبة؟

الحل: افرض X يمثل عدد الكرات البيضاء في العينة، من الواضح أن التجربة تحقق شروط التجربة فوق الهندسية حيث

$N_1 = 7, N_2 = 13, N = 20, n = 4$

ولذلك فالتوزيع الاحتمالي يكون

$$h(x; 4, 7, 13) = \frac{\binom{7}{x}\binom{13}{4-x}}{\binom{20}{4}} \; ; x = 0, 1, 2, 3, 4.$$

مثال (28) :

يتألف قسم الإحصاء في إحدى الجامعات من 10 أعضاء هيئة تدريس من بينهم 4 نساء، أريد تأليف لجنة من 3 أعضاء هيئة تدريس من هذا القسم، جد التوزيع الاحتمالي لعدد النساء في اللجنة.

الحل: افرض X يمثل عدد النساء في اللجنة.

من الواضح أن التجربة نحقق شروط التجربة فوق الهندسية حيث $N = 10$، $N_1 = 4$، $n = 3$، $N_2 = 10 - 4 = 6$

ولذلك فالتوزيع الاحتمالي يكون، حسب التعريف:

$$h(x; 3, 4, 6) = \frac{\binom{4}{x}\binom{6}{3-x}}{\binom{10}{3}} \; ; x = 0, 1, 2, 3$$

لاحظ في هذا المثال أن $N_1 > n$ أي أن حجم العينة (عدد أفراد اللجنة) أصغر من عدد النساء في القسم، ولذلك فإن ، عدد النساء في اللجنة، يأخذ القيم 3 ,2 ,1 ,0 وإذا أردت حساب القيم في هذا التوزيع فإن ذلك يكون بتطبيق القانون أعلاه، أي:

$$P(X = 0) = P(\text{لا يوجد نساء في اللجنة}) = \frac{\binom{4}{0}\binom{6}{3}}{\binom{10}{3}} = \frac{20}{120} = \frac{1}{6}$$

$$P(X=1) = P\ (\text{امرأة واحدة في اللجنة}) = \frac{\binom{4}{1}\binom{6}{2}}{\binom{10}{3}} = \frac{60}{120} = \frac{1}{2}$$

$$P(X=2) = P\ (\text{عدد النساء في اللجنة } 2) = \frac{\binom{4}{2}\binom{6}{1}}{\binom{10}{3}} = \frac{36}{120} = \frac{3}{10}$$

$$P(X=3) = P\ (\text{عدد النساء في اللجنة } 3) = \frac{\binom{4}{3}\binom{6}{0}}{\binom{10}{3}} = \frac{4}{120} = \frac{1}{30}$$

نظرية (6) :

إذا كان توزيع المتغير العشوائي المنفصل X توزيع فوق الهندسي، أي

$$h(x; n, N_1, N_2) = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}}, \quad x = 0, 1, 2, \ldots, n$$

حيث $N = N_1 + N_2$

فإن معدل X أي توقع X هو

$$M = E[X] = n \cdot \frac{N_1}{N}$$

وتباين X هو $\sigma^2 = n\frac{N_1}{N}\frac{N_2}{N}.\frac{N-n}{N-1}$

مثال (29) :صف فيه 10 طلاب، 8 طالبات، أردنا اختيار لجنة من 3، فما هو معدل عدد الطلاب في اللجنة وما هو التباين.

الحل: افرض X عدد الطلاب في اللجنة. نستعمل النظرية (6) حيث

$N_1 = 10, N_2 = 8, N = 10 + 8 = 18, n = 3$

إذاً

$$M = E[X] = 3 \times \frac{10}{18} = \frac{5}{3}$$

$$\sigma^2 = 3 \times \frac{10}{18} \times \frac{8}{18} \times \frac{18-3}{18-1} = 0.654$$

(4) توزيع بواسون Poisson Distribution

إن التجارب التي تعطينا عدد النجاحات في فترة زمنية معينة أو منطقة محددة تسمى تجارب بواسون. والفترة الزمنية يمكن أن تكون ثانية أو دقيقة أو يوما أو أسبوعا أو شهرا أو غير ذلك والمنطقة المحددة يمكن أن تكون صفحة من كتاب أو مترا مربعا من المساحة أو سم مكعباً من الحجم وغير ذلك. فعدد الزبائن الذين يدخلون إلى مكتب البريد كل خمس دقائق، وعدد السيارات التي تمر عن تقاطع طرق كل دقيقة، وعدد حوادث السيارات على طريق في الأسبوع، وعدد الجرذان في المتر المربع من حقل قمح، وعدد المكالمات الهاتفية التي تصل إلى مكتب كل عشرة دقائق، وعدد الأخطاء المطبعية في الصفحة من كتاب وعدد الزبائن الذين يدخلون البنك كل خمس دقائق، كلها أمثلة على تجارب بواسون.

إن تجربة بواسون هي كل تجربة تحقق الشروط الآتية :

1. معدل عدد النجاحات التي تحدث في فترة ز منية معينة أو منطقة محددة معلوم. وليكن λ.
2. احتمال حدوث نجاح واحد في فترة زمنية قصيرة أو منطقة صغيرة يتناسب مع طول تلك الفترة أو مساحة تلك المنطقة.
3. احتمال حدوث نجاحين أو أكثر في الفترة الزمنية القصيرة أو المنطقة الصغيرة مهمل.
4. إذا اعتبرنا عدة فترات زمنية منفصلة عن بعضها البعض فإن حدوث النجاحات في أي فترة مستقل عن حدوث النجاحات في أي فترة أخرى.

يعتبر عدد النجاحات X في تجربة بواسون متغير عشوائي بواسون. ونمثل اقترانه الاحتمالي بالاقتران $P(x;\lambda)$.

تعريف (9) : توزيع بواسون :

التوزيع الاحتمالي لمتغير عشوائي بواسون X الذي يمثل عدد النجاحات في فترة زمنية معينة أو منطقة محددة هو :

$$P(x;\lambda) = P(X = x)$$

$$= \frac{e^{\lambda}\lambda^{x}}{x!}; \quad x = 0, 1, 2, \ldots$$

حيث λ هي معدل عدد النجاحات في الفترة الزمنية المعينة أو المنطقة المحددة ، $e = 2.7182 \ldots\ldots$

عدد النجاحات هنا يعني عدد المرات التي يحدث فيها حادث معين.

مثال (30) :

تصل المكالمات الهاتفية إلى مقسم أحد المستشفيات بمعدل مكالمة واحدة في الدقيقتين.

(أ) ما احتمال وصول كل من الحوادث الآتية:

0, 1, 2, 3, 4 مكالمة في فترة أربع دقائق.

ب) ما احتمال وصول 3 مكالمات على الأقل في أربع دقائق؟

الحل: افرض X = عدد المكالمات في أربع دقائق. بما أن معدل وصول المكالمات هو مكالمة واحدة في الدقيقتين، إذن معدل وصول المكالمات في أربع دقائق يكون

$$\mu = \frac{1}{2} \times 4 = 2.$$

إن X يتبع توزيع بواسون الذي معدله $\mu = 2$ إذاً يكون

$$P(x;\mu) = P(X = x) = \frac{e^{-2}2^{x}}{x!};\ x = 0, 1, 2, \ldots$$

ويكون احتمال عدم وصول مكالمات هاتفية في أربع دقائق هو $P(0; 2)$ ويساوي

$$P(X = 0) = \frac{e^{-2}.2^{0}}{0!} = e^{-2} = 0.1353$$

وتحسب الاحتمالات الأخرى كما يلي:

$$P(X = 1) = \frac{e^{-2}.2^{1}}{1!} = 2e^{-2} = 0.2706$$

$$P(X = 2) = \frac{e^{-2}.2^{2}}{2!} = \frac{4}{2}\ e^{-2} = 0.2706$$

$$P(X = 3) = \frac{e^{-2}.2^{3}}{3!} = \frac{8}{6} e^{-2} = 0.1804$$

$$P(X = 4) = \frac{e^{-2}.2^{4}}{4!} = \frac{16}{24} \times e^{-2} = 0.0902$$

وهكذا.

(ب) احتمال وصول 3 مكالمات على الأقل في أربع دقائق هو $P(X \geq 3)$ وباستعمال احتمال المتممة فإن :

$$P(X \geq 3) = 1 - P(X < 3)$$

ولكن $P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2)$

وهذه جميعها حسبت في فرع أ.

إذن $$P(X < 3) = e^{-2} + 2e^{-2} + \frac{4}{2} e^{-2}$$

$$e^{-2} (1 + 2 + 2) = 5e^{-2} = 0.6765$$

إذن $$P(X \geq 3) = 1 - 0.6765 = 0.3235$$

نظرية (7) :

إذا كان X متغير عشوائي بواسون الذي توزيعه الاحتمالي $P(x; \lambda)$ حيث λ معدل عدد الحوادث في الفترة الزمنية المعينة، فإن توقع × هو $E(X) = \lambda$، وتباين X هو $\sigma^2_x = \lambda$ والانحراف المعياري له يساوي $\sqrt{\lambda}$

العلاقة بين توزيع بواسون وتوزيع ذات الحدين.

في كثير من الأحيان يكون حساب الاحتمالات في توزيع ذات الحدين مضنيا، فعلى سبيل المثال، إذا كانت نسبة المصابين بمرض معين في بلد ما هي 0.001، فما احتمال وجود 10 مصابين بهذا المرض في حي يقطنه 5000 نسمة من ذلك البلد؟

يظهر لك بسهولة أنك لو افترضت X تساوي عدد المصابين بهذا المرض في الحي المذكور فإن توزيع X يكون ذات الحدين حيث $p = 0.001$، $n = 5000$ والمطلوب إيجاد $P(X = 10)$ أي أن:

$$P(X = 10) = b(10; 5000, 0.001) = \binom{5000}{10} (0.001)^{10} (0.999)^{5000 - 10}$$

أي أن $$P(X = 10) = \binom{5000}{10} (0.001)^{10}(0.999)^{4990}$$

وستجد صعوبة في حساب هذا المقدار، ولكن يمكن تقريب هذا الاحتمال بواسطة توزيع بواسون وذلك عندما يكون n كبيرا، p صغيرة بشكل مناسب بحيث تبقى np معتدلة القيمة، ويتضح ذلك من المثال التالي:

مثال (31) :

اكتب قيم الاحتمالات، في توزيع ذات الحدين، حيث p = 0.2، n = 20، وقارنها بالاحتمالات التي تحصل عليها باستعمال توزيع بواسون حيث $\lambda = np$.

الحل: (أ) باستعمال توزيع ذات الحدين

$$P(X = 0) = \binom{20}{0}(0.2)^0\,(0.8)^{20} = 0.012$$

$$P(X = 1) = \binom{20}{1}(0.2)^1\,(0.8)^{19} = 0.057$$

$$P(X = 2) = \binom{20}{2}(0.2)^2\,(0.8)^{18} = 0.137$$

(ب) باستعمال توزيع بواسون $P(x; \lambda) = \dfrac{e^{-\lambda}\lambda^x}{x!}$

حيث $= 4\lambda = np = 20 \times 0.2$

فنجد

$$P(X = 0) = \frac{e^{-4}\,4^0}{0!} = 0.0183$$

$$P(X = 1) = \frac{e^{-4}\,4^1}{1!} = 0.0732$$

$$P(X = 0) = \frac{e^{-4}\,4^2}{2!} = 0.01464$$

وهكذا.

وعندما نقارن الإجابات نرى أن تقريب ذات الحدين بتوزيع بواسون تقريب جيد إلى حدّ ما عندما تكون $n = 20$ وهو يتحسن عندما تكبر قيمة n، وبالتالي يمكن إعطاء القاعدة التالية:

قاعدة (1): إذا كانت n كبيرة و p صغيرة بحيث تبقى np معتدلة القيمة فإن توزيع ذات الحدين $b(x; n, p)$ يمكن تقريبه بتوزيع بواسون الذي معدله $\lambda = np$ وفي هذه الحالة تكتب: $b(x; n,p) \approx P(x; \lambda)$ حيث $\approx$ تعني تساوي تقريبا.

في كثير من الأحيان إذا كانت $n \geq 100$، $np \geq 10$ فإن التقريب يكون ممتازا وبإمكانك أيضا الاعتماد على هذا التقريب إذا كانت n كبيرة، np في حدود 5.

مثال (32) :

تألف امتحان اللغة الإنجليزية كلغة أجنبية من 100 سؤال، لكل منها 4 إجابات، واحدة فقط منها صحيحة. أجاب أحد الطلبة عن جميع الأسئلة بالتخمين.

ما احتمال أن يحصل الطالب على 10 إجابات صحيحة؟

الحل: احتمال الحصول على إجابة صحيحة في أي سؤال يساوي $\frac{1}{4}$، وهناك 100 سؤال، وبالتالي إذا فرضت أن X تساوي عدد الإجابات الصحيحة فإن توزيع X هو ذات الحدين فيه

$$n = 100, \; p = \frac{1}{4}$$

احتمال حصول الطالب على 10 إجابات صحيحة =

$$P(X = 10) = b(10; 100, \frac{1}{4}) = \binom{100}{10}\left(\frac{1}{4}\right)^{10}\left(\frac{3}{4}\right)^{90}$$

وحساب هذا المقدار يتطلب جهداً كبيراً ولذلك نعمد إلى تقريب الجواب بواسطة توزيع بواسون الذي فيه $\lambda = np$ أي أن $\lambda = 100 \times \frac{1}{4} = 25$

فيكون:

$$P(X = 10) = \frac{e^{-25}(25)^{10}}{10!}$$

(5) التوزيع الهندسي Geometric Distribution

هناك بعض التجارب الشبيهة بتجربة ذات الحدين ولكن عدد المحاولات لا يكون فيها محددا في البداية. عد إلى تجربة ذات الحدين وافرض أن التجربة تحقق الشروط (1) ,(2) ,(3) في ذلك التعريف، ولا تحقق الشرط (4) وبدلا من إجراء التجربة n من المحاولات، افرض أنك تستمر في إجراء التجربة حتى تحصل على أول نجاح، وعندها تتوقف.

الآن، افرض X تمثل عدد المحاولات التي احتجت إلى القيام بها للحصول على أول نجاح، يسمى X متغيرا عشوائيا هندسيا.

مثال (33) :

قم برمي زهرة نرد منتظمة حتى يظهر العدد 3 إلى أعلى. من الواضح أن احتمال ظهور 3 إلى أعلى هو $p = \frac{1}{6}$

افرض X يمثل عدد الرميات التي تحتاجها للحصول على 3، فتلاحظ أن القيم التي يأخذها X هي 1, 2, 3, 4, ...

وذلك لأنه إذا ظهرت 3 من أول رمية فيكون X = 1 وإذا لم تظهر 3 في المرة الأولى وظهرت في المرة الثانية فإن X = 2 وهكذا.

تعريف (9) :

إذا حققت تجربة ما الشروط (2) ,(3),(1) من تعريف تجربة ذات الحدين، فإن X الذي يمثل عدد المحاولات التي تحتاجها للحصول على أول نجاح يسمى متغيرا عشوائيا هندسيا ويسمى توزيعه الاحتمالي التوزيع الهندسي، ويعطى بالمعادلة:

$$P(X = x) = g(x; p) = q^{x-1} p \quad , \quad x = 1, 2, 3, ...$$

حيث p احتمال النجاح في الرمية الواحدة، احتمال "الفشل" ويساوي q = 1-p

مثال (34) :

قم برمي زهرة نرد منتظمة حتى يظهر العدد 1 إلى أعلى. ما احتمال أنك تحتاج إلى 3 محاولات.

الحل: X =عدد المحاولات حتى تحصل على1، احتمال النجاح $p = \frac{1}{6}$ إذاً $q = \frac{5}{6}$

وبالتالي:

$$P(X = 3) = \left(\frac{5}{6}\right)^{3-1} \cdot \frac{1}{6} = \frac{25}{216} = 0.1157$$

مثال (35) :

رميت قطعة نقود متزنة حتى حصلت على صورة H. ما احتمال أنك احتجت إلى 4 رميات على الأقل؟

الحل: احتمال الحصول على الصورة H في الرمية الواحدة يساوي $\frac{1}{2}$ أي أن $q = \frac{1}{2}, p = \frac{1}{2}$

افرض X عدد المحاولات التي تحتاجها للحصول على الصورة، فيكون المطلوب $P(X \geq 4)$

لاحظ أن X متغير عشوائي هندسي وباستعمال قانون المتممة تجد:

$$P(X \geq 4) = 1 - P(X < 4)$$

$$P(X < 4) = P(X = 1) + P(X = 2) + P(X = 3)$$

$$= \left(\frac{1}{2}\right)^{1-1} \cdot \frac{1}{2} + \left(\frac{1}{2}\right)^{2-1} \times \frac{1}{2} + \left(\frac{1}{2}\right)^{3-1} \cdot \frac{1}{2}$$

$$= \frac{1}{2} + \frac{1}{4} + \frac{1}{8} = \frac{7}{8}$$

إذاً $P(X \geq 4) = 1 - \frac{7}{8} = \frac{1}{8}$

نظرية (8) :

إذا كان X متغيراً عشوائياً توزيعه التوزيع الهندسي

$$P(X = x) = g(x; p) = q^{x-1} p; x = 1, 2, 3, ...$$

حيث p احتمال "النجاح" في الرمية الواحدة، q = 1 - p احتمال "الفشل" في الرمية الواحدة، فإن معدل X أي توقع X يكون

$$\mu = E[X] = \frac{1}{p}$$

وتباين X يكون

$$\sigma^2 = \frac{q}{p^2}$$

مثال (36) : إذا قمت بالتجربة التالية:

اسحب ورقة من مجموعة ورق اللعب المكوّنة من 52 ورقة ولاحظ نوعها ثم أعدها للمجموعة. كرّر هذه المحاولة لحين الحصول على أول صورة(يعني Q, أو J أو K)

(1) ما معدل عدد المحاولات التي تحتاجها؟

(2) ما احتمال أنك تحتاج إلى 3 محاولات أو أقل؟

الحل: من الواضح أن هذه التجربة تحقق الشروط الثلاثة الأولى لتجربة ذات الحدين. افرض X يمثل عدد المحاولات المطلوبة للحصول على أول صورة، فيكون توزيع X التوزيع الهندسي الذي فيه

$p = \frac{3}{13} = \frac{12}{52}$، $q = 1 - p = \frac{10}{13}$.

وتوزيع X هو

$$P(X = x) = \left(\frac{10}{13}\right)^{x-1} \cdot \frac{3}{13}, x = 1, 2, ...$$

(1) من النظرية (8) نجد :

$$E[X] = \frac{1}{p} = \frac{1}{\frac{3}{13}} = \frac{13}{3} = 4\frac{1}{3}$$

معدل عدد المحاولات المطلوبة التي تحتاجها للحصول على أول صورة.

(2) احتمال الحاجة إلى 3 محاولات أو أقل هو

$P(X \le 3) = P(X = 1) + P(X = 2) + P(X = 3)$

$$= \left(\frac{10}{13}\right)^{1-1}\frac{3}{13} + \left(\frac{10}{13}\right)^{2-1}\frac{3}{13} + \left(\frac{10}{13}\right)^{3-1}\frac{3}{13}$$

$$= \frac{3}{13}\left[1 + \frac{10}{13} + \frac{10^2}{13^2}\right] = \frac{3}{13} \times \frac{399}{169} = 0.545$$

5 : 6 المتغيرات العشوائية الثنائية Bivariate random variables

لقد اقتصرت دراستنا حتى الآن على المتغيرات العشوائية باعتبار كل منها على حده، مثل العدد الظاهر على زهرة نرد عند رميها مرة واحدة، أو عدد الحوادث في الأسبوع عند تقاطع طرق محدد، أو عدد أفراد عائلة اختيرت عشوائيا وهكذا.

ولكن هناك دراسات هامة مثل التنبؤ أو الارتباط أو الدراسات النظرية في الإحصاء الاستقرائي، نحتاج فيها إلى اعتبار متغيرين اثنين أو أ كثر في نفس الوقت، فمثلا ربما يرغب تاجر في تنبؤ مبيعاته (المتغير Y) بناء على مصروفاته في الدعاية (المتغير X). أو ربما يريد أحد المختصين في التربية معرفة العلاقة بين دخل الأسرة (المتغير X) ومؤهل رب العائلة الأكاديمي (المتغير Y).

عندما نأخذ بالاعتبار متغيرين أو أكثر في نفس الوقت فإن البناء الاحتمالي يعرف بدلالة التوزيعات الثنائية أو المتعددة.

تعريف (10) :

اقتران الاحتمال المشترك لمتغيرين عشوائيين منفصلين X, Y هو الاقتران :

$f(x, y) = P(X= x , Y = y)$

لجميع القيم x, y للمتغيرات العشوائية X, Y.

أي أنه :

إذا كان (X, Y) متغيرا عشوائيا ثنائيا، مركبتا X, Y منفصلتان وكان f(x, y) اقتراناً بحيث :

(1) $f(x, y) \geq 0$ لجميع القيم x, y

(2) $\sum_{\text{جميع}(x,y)}\sum \mathbf{f(x,y) = 1}$ فإن الاقتران f(x, y) يسمى اقتران الكثافة المشتركة (الاقتران الاحتمالي المشترك) للمتغير الثنائي (X, Y).

تعريف (11) : يسمى $\mathbf{f_1(x) = \sum_y f(x,y)}$ اقتران الكثافة الهامشية (الحديّة) للمتغير X (marginal density).

يسمى $\mathbf{f_2(y) = \sum_x f(x,y)}$ اقتران الكثافة الهامشية للمتغير Y.

وكما هو الحال في المتغيرات العشوائية المنفصلة الأحادية فإن الكثافة الاحتمالية المشتركة للمتغير العشوائي المنفصل الثنائي (X, Y) يمكن التعبير عنه بجدول ثنائي يعطي قيم X في الصف الأفقي وقيم Y في السطر العمودي، وقيم الاحتمالات المشتركة في خلال الجدول.

مثال (37) : يعطي الجدول التالي اقتران الكثافة المشتركة للمتغيرين X , Y حيث X عدد المكالمات المحلية، Y عدد المكالمات الوطنية التي يستقبلها مقسم خلال عشرة دقائق.

Y \ X	0	1	2	3	4	5
0	0.20	0.11	0.10	0.07	0.05	0.01
1	0.15	0.06	0.05	0.06	0.04	0.00
2	0.04	0.03	0.02	0.00	0.01	0.00

أ) أوجد $P(X = 0 , Y = 2)$

الحل : من الجدول مباشرة ، الجواب 0.04

ب) أوجد $P(X \leq 2 , Y = 1)$

من الواضح أن :

$P(X \le 2, Y = 1) = P(X = 0, Y = 1) + P(X = 1, Y = 1) + P(X = 2, Y = 1)$

$= 0.15 + 0.06 + 0.05 = 0.26$

جـ) أوجد $P(X = 0)$

من الواضح أن $P(X = 0) = 0.02 + 0.15 + 0.04 = 0.39$

د) أوجد $P(X = 3 \mid Y = 0)$

الحل :

$$P(X = 3 \mid Y = 0) = \frac{P(X = 3, Y = 0)}{P(Y = 0)}$$

$$= \frac{0.07}{0.20 + 0.11 + 0.10 + 0.07 + 0.05 + 0.01}$$

$$= \frac{0.07}{0.54} = \frac{7}{54}$$

هـ) أوجد اقتران الكثافة الهامشية للمتغير X .

الحل :

$$f_1(x) = \sum_y f(x, y)$$

ولإيجاد $f_1(x)$ نثبت x في كل مرة ونجمع الخلايا المقابلة لها لجميع قيم Y ونضع النتيجة في جدول :

x	0	1	2	3	4	5	**المجموع**
$f_1(x)$	0.39	0.20	0.17	0.13	0.10	0.01	1

أي أننا نجمع عموديا فوق y فنجد $f_1(x)$.

(و) أوجد اقتران الكثافة الهامشية للمتغير Y.

الحل : $f_2(y) = \sum_x f(x, y)$

ونجد ذلك بجمع الخلايا أفقيا على x، فنحصل على :

y	$f_2(y)$
0	0.54
1	0.36
2	0.10
المجموع	1

مـن المفـاهيم الهامــة التــي درسـنـاها ســابقا مفهـوم الاسـتقلال بـين الحوادث ويمكننا تعميم ذلك المفهوم لدراسة استقلال المتغيرات العشوائية.

تعريف (12) :

يكون المتغيران Y , X مستقلين Independent إذا وفقط إذا كان $f(x,y) = f_1(x).\ f_2(y)$ لكل y , x.

مثال (38) : يمثل الجدول التالي اقتران الكثافة المشتركة للمتغير (X, Y) وقد حسبنا منه اقتراني الكثافة الهامشية $f_1(x)$، $f_2(y)$. هل Y, X مستقلان ؟.

X / Y	1	2	3	4	$f_2(y)$
3	0.06	0.03	0.09	0.12	0.3
5	0.04	0.02	0.06	0.08	0.2
10	0.10	0.05	0.15	0.20	0.5
$f_1(x)$	0.2	0.1	0.3	0.4	1

لاحظ أن $f(x, y) = f_1(x)f_2(y)$ لجميع y, x فمثلا

$$f(1, 3) = 0.06 = 0.2 \times 0.3 = f_1(1).f_2(3)$$

أي أن القيمة في كل خلية هي حاصل ضرب القيمة الهامشية في صفها مع القيمة الهامشية في عمودها.

إذاً Y, X مستقلان.

لاحظ أنه لبرهان الاستقلال يجب أن يتحقق التعريف لجميع القيم y, x أما إذا كان Y, X غير مستقلين فيكفي لبرهان ذلك أن خلية واحدة على الأقل لا تساوي حاصل ضرب صفها مع عمودها.

مثال (39) :

يعطي الجدول التالي الكثافة الاحتمالية المشتركة للمتغيرين Y, X هل Y, X مستقلان :

X / Y	1	2	$f_2(y)$
1	0.1	0.0	0.1
4	0.1	0.3	0.4
5	0.3	0.2	0.5
$f_1(x)$	0.5	0.5	1

الحل : نجد اقتراني الكثافة الهامشية لكل من Y, X أي $f_1(x)$, $f_2(y)$ لاحظ الخلية $f(1, 1) = 0.1$ بينما صفها 0.1 وعمودها 0.5 أي $0.1 \neq 0.5 \times 0.1$ أي $f(1,1) \neq f_1(1) . f_2(1)$ إذا Y, X غير مستقلين.

لو أخذت خلية أخرى مثلا :

$$f(2,1) = 0 \neq 0.5 \times 0.1$$

إذا Y, X غير مستقلين (Not independent).

تعريف (13) :

إذا كان (X, Y) متغيرا عشوائيا ثنائيا اقتران كثافته المشتركة f (x, y) وكان g(X, Y) اقترانا في Y, X فإن توقع g (X, Y) يعرف كما يلي :

$$E[g(X,Y)] = \sum_y \sum_x g(x,y) f(x,y)$$

شريطة أن حاصل الجمع موجود أي إذا كان حاصل الجمع عبارة عن متسلسلة فيجب أن تكون متقاربة ومن الحالات الخاصة التي نحتاج إلى

دراستها هي :

$g(X, Y) = X$ فيكون التوقع هو $E[X]$

$g(X, Y) = Y$ فيكون التوقع هو $E[Y]$

$g(X, Y) = XY$ فيكون التوقع هو العزم المشترك $E[XY]$.

مثال (40) : يعطي اقتران الكثافة المشتركة للمتغيرين (X, Y) بالجدول التالي :

X / Y	0	1	2	$f_2(y)$
1	$\frac{1}{12}$	$\frac{3}{12}$	$\frac{1}{12}$	$\frac{5}{12}$
2	$\frac{4}{12}$	0	$\frac{3}{12}$	$\frac{7}{12}$
$f_1(x)$	$\frac{5}{12}$	$\frac{3}{12}$	$\frac{4}{12}$	1

أوجد :

$E[X\ Y]\ ,\ E[Y]\ ,\ E[X]$

الحل :

$$E[X] = \sum_y \sum_x xf(x, y)$$

$$= 0 \times f(0,1) + 0.f(0,2) + 1\,f(1,1) + 1\,f(1,2) + 2f(2,1) + 2f(2,2)$$

$$= 0 + 0 + \frac{3}{12} + 0 + 2 \times \frac{1}{12} + 2 \times \frac{3}{12} = \frac{11}{12}$$

وبالمثل :

$$E[Y]=1\times\frac{1}{12}+1\times\frac{3}{12}+1\times\frac{1}{12}+2\times\frac{4}{12}+2\times 0+2\times\frac{3}{12}=\frac{19}{12}$$

$$E[XY]=0\times 1\times\frac{1}{12}+0\times 2\times\frac{4}{12}+1\times 1\times\frac{3}{12}$$
$$+1\times 2\times 0+2\times 1\times\frac{1}{12}+2\times 2\times\frac{3}{12}$$
$$=\frac{17}{12}$$

لاحظ أنه إذا كان g (X,Y) اقترانا بدلالة متغير واحد فقط مثل X, X^2 أو Y, Y^3 وهكذا فيمكننا حساب التوقع الرياضي بأن نجد أولا اقتران الكثافة الهامشية للمتغير ثم حساب التوقع الرياضي المطلوب، ففي مثالنا السابق، نضع اقتراني الكثافة الهامشية لكل من Y, X على الجدول نفسه ونحسب E[X] , E[Y] منهما مباشرة، فنحصل على:

$$\mu_x=E[X]=0\times\frac{5}{12}+1\times\frac{3}{12}+2\times\frac{4}{12}=\frac{11}{12}$$
$$\mu_y=E[Y]=1\times\frac{5}{12}+2\times\frac{7}{12}=\frac{19}{12}$$

تعريف (14) :

إذا كان (X, Y) متغيرا عشوائيا ثنائيا اقتران كثافته المشتركة f (x, y) فإن التباين المشترك بين X, Y (Covariance). يعرف كما يلي :

$$Cov(X,Y)=E[(X-\mu_x)(Y-\mu_y)]$$
حيث
$$\mu_y=E[Y],\ \mu_x=E[X]$$

ويمكننا بسهولة برهنة أن :

$$Cov(X,Y)=E[XY]-\mu_x\mu_y$$

مثال (41) :

بالرجوع إلى المثال السابق، أوجد التباين المشترك بين Y, X.

الحل :

$$Cov(X,Y) = E[XY] - \mu_x \mu_y$$

$$= \frac{17}{12} - \frac{11}{12} \times \frac{19}{12} = \frac{204-209}{144} = \frac{-5}{144}$$

تعريف (15) :

إذا كان (X ; Y) متغيرا عشوائيا ثنائيا فإن معامل الارتباط بين Y, X يعرف كما يلي :

$$\rho = \rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

حيث : σ_x الانحراف المعياري للمتغير X ،

σ_y الانحراف المعياري للمتغير Y.

مع ملاحظة أن : $\sigma_X^2 = E[X^2] - \mu_x^2$

$$\sigma_Y^2 = E[Y^2] - \mu_y^2$$

مثال (42) :

احسب معامل الارتباط بين Y, X في المثال السابق.

الحل : نحتاج إلى حساب $E[X^2]$, $E[Y^2]$ حيث المقادير الأخرى محسوبة :

$$E[X^2] = 0^2 \times \frac{5}{12} + 1^2 \times \frac{3}{12} + 2^2 \times \frac{4}{12} = \frac{19}{12}$$

$$E[Y^2] = 1^2 \times \frac{5}{12} + 2^2 \times \frac{7}{12} = \frac{33}{12}$$

إذا :

$$\sigma_X^2 = \frac{19}{12} - \left(\frac{11}{12}\right)^2 = \frac{228-121}{144} = \frac{107}{144}$$

$$\sigma_Y^2 = \frac{33}{12} - \left(\frac{19}{12}\right)^2 = \frac{396-361}{144} = \frac{35}{144}$$

وبالتعويض في تعريف معامل الارتباط نجد :

$$\rho = \frac{-5/144}{\sqrt{\frac{107}{144}}\sqrt{\frac{35}{144}}} = \frac{-5}{10.34 \times 5.92} = \frac{-5}{61.17} = -0.08$$

نظرية (9) :

(أ) إذا كان Y, X مستقلين فإن :

$$E[XY] = E[X] . E[Y]$$

ولكن العكس ليس بالضرورة صحيحا.

(ب) إذا كان Y, X مستقلين فإن

$$Cov(X, Y) = 0$$

والعكس غير صحيح بالضرورة.

(جـ) إذا كان Y, X مستقلين فإن :

$$\rho(X, Y) = 0$$

والعكس غير صحيح بالضرورة، أي أنه يمكن أن يكون معامل الارتباط بين Y, X صفرا ولكن Y, X غير مستقلين.

(د) إذا كان ρ معامل الارتباط بين Y, X فإن

$$-1 \leq \rho(X, Y) \leq 1$$

نظرية (10) :

إذا كان (X, Y) متغيرا عشوائيا اقتران كثافته f (x, y) فإن :

أ- $Var(X \pm Y) = Var(X) + Var(Y) \pm 2\, Cov(X, Y)$

ب- $Cov(aX, bY) = ab\, Cov(X, Y)$

مثال (43) :

إذا كان (X, Y) متغيرا عشوائيا ثنائيا بحيث $Va(X) = 16$, $Var(Y) = 25$, $Cov(X, Y) = 15$.

أوجد :

أ) $Var(X - Y)$ ب) $Var(X + Y)$

جـ) $Var(3X - 4Y)$ د) $Cov(2X, 5Y)$

الحل :

أ- $Var(X - Y) = Var(X) + Var(Y) - 2Cov(X,Y)$

$= 16 + 25 - 2 \times 15 = 11$

ب- $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X,Y)$

$= 16 + 25 + 2 \times 15 = 71$

جـ- $Var(3X - 4Y) = Var(3X) + Var(4Y) - 2\, Cov(3X, 4Y)$

$= 3^2 \times 16 + 4^2 \times 25 - 2 \times 3 \times 4 \times 15$

$= 144 + 400 - 360 = 184$

د- $Cov(2X, 5Y) = 2 \times 5 \times Cov(X, Y) = 2 \times 5 \times 15 = 150$

تمارين:

1-5 : يمثل الجدول التالي التوزيع الاحتمالي لعدد الأعطال الأسبوعية في مركز الحاسوب.

x	f(x)
0	0.10
1	0.05
2	0.15
3	0.25
4	0.20
5	0.15
6	0.10

أ. ما احتمال حدوث أكثر من عطلين في أسبوع معين.

ب. إذا علم أنه حدثت بعض الأعطال في أسبوع معين فما احتمال أن عدد الأعطال كان أربعة.

جـ. كم عطلاً تتوقع أن يحدث في أسبوع ما؟

2-5 : رميت زهرة نرد مرتين وسجّل العدد الظاهر إلى أعلى في كل مرّة. افرض X يمثل القيمة المطلقة للفرق بين العددين الظاهرين. أوجد التوزيع الاحتمالي للمتغير X وأوجد توقع وتباين X.

3-5 : يمثل الجدول التالي التوزيع الاحتمالي للمتغير X حيث $E[X] = 2.8$ أوجد b, a واحسب تباين X.

x	0	2	3	4	5
f(x)	0.1	a	0.2	0.1	b

4-5 : إذا كان X متغيراً عشوائياً منفصلاً له $\sigma_x^2 = 4$, $\mu_x = 10$. افرض $Y = 0.3X - 7$ أوجد $E[Y]$، تباين Y.

5-5 : X متغير عشوائي توزيعه الاحتمالي كما في الجدول التالي:

x	-2	0	2	3
f(x)	0.30	0.10	0.25	0.35

أوجد التوقع الرياضي والتباين للمتغير X.

أوجد $E[X^2]$. أوجد $E[2X + 11]$.

6-5 : الاقتران= $f(x) = \binom{3}{x}(0.2)^x(0.8)^{3-x}$ حيث x = 0, 1, 2, 3 يمثل الاقتران الاحتمالي للمتغير X.

احسب توقع X وتباين X.

7-5 : أراد أحد المستثمرين استثمار بعض المال في شركة مساهمة فعلم أنه سيربح مبلغ 5000 ديناراً باحتمال 0.4 وأنه يخسر 1000 ديناراً باحتمال 0.6. ما توقع الربح لدى هذا المستثمر؟

8-5 : يوجد في معرض تلفزيونات 12 جهازاً من ضمنها جهازان فيهما عطل. اشترى أحد التجار 3 أجهزة. افرض X عدد الأجهزة التي فيها عطل.

أوجد توقع X وتباين X بدون استعمال نظريات التوزيع فوق هندسي.

9-5 : يسجل لاعب كرة سلّة بالمعدل 7 أهداف في كل عشرة محاولات. ما احتمال أن هذا اللاعب يسجل 12 هدفاً إذا حاول رمي الكرة 15 مرة، وكم هدفاً يتوقع أن يسجل في هذه المحاولات؟

10-5 : لدى بستاني 15 من أبصال الزهور مؤلفة من 5 بصلات أزهار (أ)، 6 أبصال أزهار (ب)، 4 بصلات أزهار (جـ). أخذ البستاني 7 بصلات وزرعها، ما احتمال أن تكون هذه البصلات مؤلفة من 2 نوع (أ) 3 نوع (ب)، 2 نوع (جـ).

إذا كان X عدد البصلات من نوع (جـ) في العيّنة التي زرعها البستاني،

اكتب التوزيع الاحتمالي للمتغير X.

ما التوقع الرياضي والتباين للمتغير X؟

11-5 : إذا كان احتمال أن تعلق السمكة بسنارة أحد الصيادين 0.3 في الرمية الواحدة، ما احتمال أن يصيد هذا الصياد أول سمكة في الرمية الخامسة؟ ما احتمال أن يفشل الصياد ست مرّات قبل أن يصيد أول سمكة.

ما عدد المرات التي يتوقع هذا الصياد محاولتها لاصطياد سمكة واحدة؟

12-5 : إذا كان (X, Y) متغيرا عشوائيا ثنائيا اقتران كثافته الاحتمالية معطى في الجدول التالي :

Y \ X	1	2	3
0	$\frac{1}{9}$	$\frac{1}{18}$	$\frac{1}{6}$
1	$\frac{1}{9}$	$\frac{1}{6}$	$\frac{1}{18}$
4	$\frac{1}{18}$	$\frac{1}{6}$	$\frac{1}{9}$

أ- أوجد $P\ (X > 2, Y > 1)$

ب- أوجد $P\ (X = 3 \mid Y = 4)$

جـ- أوجد $P\ (X \le 2\ , Y = 1)$

د - أوجد $P\ (X = 1\ \ , Y = 1)$

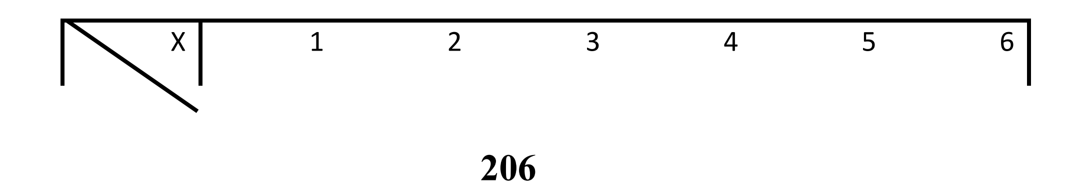
13-5 : في تمرين (12) أعلاه أوجد الكثافة الاحتمالية الهامشية لكل من Y, X.

هل Y, X مستقلان ؟

14-5 : في التمرينين (12), (13) أعلاه :

أوجد $\mu_X,\ \mu_Y,\ \sigma_X^2,\ \sigma_Y^2,\ Cov(X, Y)$ ، معامل الارتباط بين Y, X.

15-5 : في تجربة رمي زهرة نرد ثم رمي قطعة نقود، إذا كان X يمثل العدد الظاهر على زهرة النرد، Y يمثل الوجه H أو الكتابة T التي تظهر على قطعة النقود. املأ الجدول التالي :



X	1	2	3	4	5	6

	Y
	H
	T

أوجد E [XY] هل X ,Y مستقلان ؟

الفصل السادس
التوزيعات الاحتمالية المتصلة
Continuous Probability Distributions

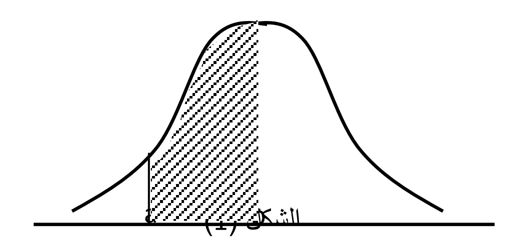
6 : 1 مقدّمة

يصادفنا في كثير من الأحيان متغيرات عشوائية تأخذ جميع القيم في فترة ما . مثل هذه المتغيرات تسمى متغيرات عشوائية متصلة. ولما كان المتغير العشوائي المتصل يأخذ جميع القيم في فترة ما فهذا يعني أنك لا تستطيع البدء في عد تلك القيم ولكنك تقيسه بشكل تقريبي، فمثلا، طول الطالب في عمر 16 سنة متغير عشوائي متصل حيث أنه يمكن أن يأخذ أي قيمة في فترة ما مثل الفترة من 150.0 سم إلى 190.0 سم. ولو أخذنا أي فترة ولتكن 170.5سم إلى 171.5سم لوجدنا آلاف الطلبة الذي تقع أطوالهم داخلها، وبالتالي فاحتمال أن يكون طول أحد الطلبة قيمة محددة واحدة مثل 165.3سم سيكون صفرا. لأن الاحتمال في حالة المتغير العشوائي المتصل عبارة عن المساحة تحت منحنى التوزيع الاحتمالي بين أي نقطتين معينتين. ومن المعروف أن المساحة فوق نقطة واحدة تساوي صفرا. هذا إضافة إلى أنه حتى في تفسيرنا الاحتمال على أنه نهاية التكرار النسبي فإن هذا التكرار النسبي للطلبة الذين أطوالهم 165.3سم بالضبط يساوي صفرا. لأنه مهما صغرت الفترة حول القيمة يظل هناك أعداد كبيرة جدا من الطلبة الذين تقع أطوالهم في تلك الفترة، وبالتالي تكون نسبة الطلبة الذين طولهم 165.3 بالنسبة للطلبة الذين تقع أطوالهم في الفترة حول 165.3 تساوي صفرا.

وبالتالي فإن احتمال أي قيمة معينة بحد ذاتها سيكون صفرا. وهذا يعني أن من صفات المتغير العشوائي المتصل أن احتمال مساواته لأي قيمة معينة يكون صفرا. ولهذا لا يمكن تمثيل التوزيع الاحتمالي له بجدول، فنعبر عنه بمعادلة، ويكون الاحتمال عبارة عن مساحة تحت منحنى المعادلة، وليس بجمع احتمال النقاط لأن احتمال كل نقطة يساوي صفرا.

تعريف (1) :

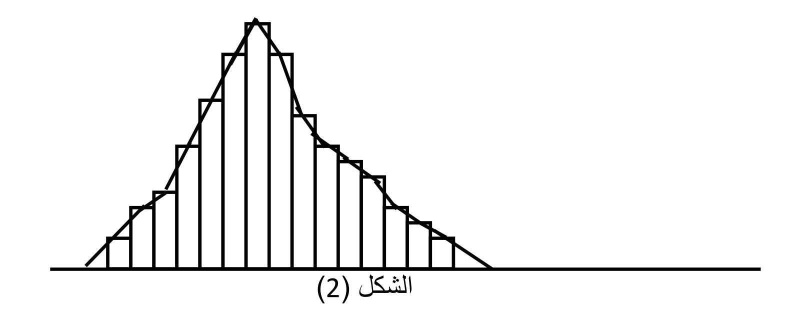
إذا كان X متغيرا عشوائيا متصلا وكان f(x) اقترانا حقيقيا غير سالب بحيث تكون المساحة تحته تساوي 1 فإن f(x) يسمى اقتران الكثافة الاحتمالية أي التوزيع الاحتمالي المتصل للمتغيرة X. أما احتمال وقوع X بين قيمتين X = b , X = a فيساوي المساحة تحت منحنى f(x) وفوق المحور الأفقي والمحصورة بين b, a. أنظر الجزء المظلل في الشكل (1).



الشكل (1)

ولما كان الاحتمال للمتغير العشوائي المتصل عبارة عن مساحة فإننا نجد ذلك الاحتمال بإيجاد المساحة عن طريق النظريات الهندسية أو التكامل أو باستعمال جداول خاصة لبعض الاقترانات الهامة كما سنوضحه مستقبلا.

ويمكن الحصول على منحنى التوزيع الاحتمالي للمتغير العشوائي المتصل عمليا بطريقة الحصول على منحنى التوزيع التكراري النسبي ذي الفئات، وذلك بأن ترسم المضلع التكراري النسبي ثم تحاول تمهيد ذلك المضلع برسم المنحنى الذي يمكن أن يطابقه. كما يظهر في الشكل (2) حيث نجد الخطوط المنكسرة تمثل المضلع التكراري النسبي وقد تم تمهيدها بتطبيق منحنى عليها بشكل تقريبي.



الشكل (2)

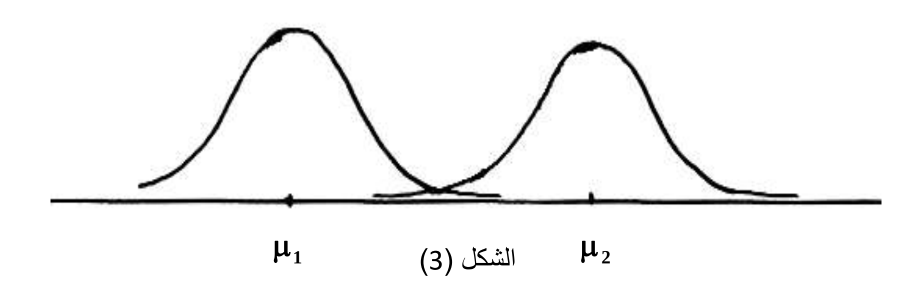
6 : 2 : التوزيع الطبيعي The Normal Distribution :

يعتبر التوزيع الطبيعي من أهم التوزيعات الاحتمالية المتصلة وتأتي أهميته من الناحيتين النظرية والتطبيقية.

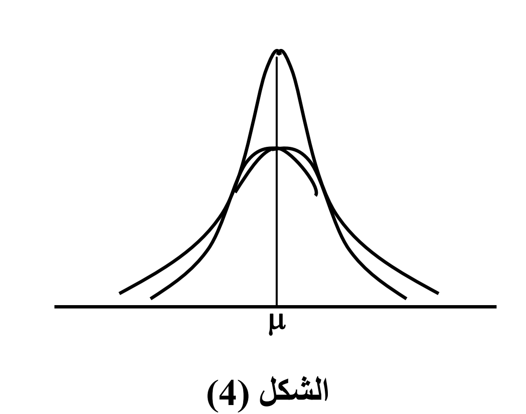
يوصف التوزيع الطبيعي بمعادلة رياضية تحدّد منحناه، وهي تتعين تماما بمعرفة كل من المعدل μ والتباين σ^2أي أنك إذا علمت أن متغيرا عشوائيا يخضع لتوزيع طبيعي فإن هذا التوزيع يتحدد تماما بمعرفة معدل ذلك التوزيع μ وتباينه σ^2. وبالتالي فأنت تستطيع رسم منحناه الذي يشبه شكل الجرس، وهو متماثل حول الخط المستقيم $x = \mu$، ويتقارب من الصفر على الجهتين عندما $x \to -\infty$ وعندما $x \to \infty$ (أي عندما تؤول x إلى سالب مالا نهاية وكذلك عندما تؤول x إلى موجب مالا نهاية).

وتعين μ معدل التوزيع، وهي مركز التوزيع، فلو أقيم عليها عمود فإن التوزيع يكون متماثلا حوله، أما σ فتعين الانحراف المعياري (أحد مقاييس التشتت)، ولو تحركت μ إلى اليمين أو اليسار لانتقل مركز التوزيع تبعا لذلك ولا يتغير شكل المنحنى، أما إذا تغيرت σ وبقيت μ نفسها دون تغير فإن مركز التوزيع لا يتغير، لكن تشتت التوزيع يقل كلما صغرت σ ويكبر كلما كبرت، أي أن تباعد المنحنى يقل كلما صغرت σ ويزيد كلما كبرت.

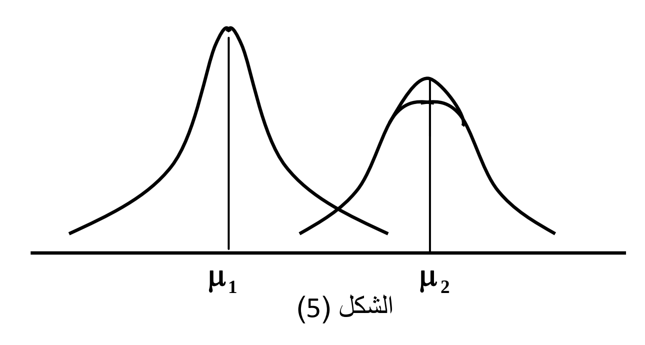
أما إذا تغيرت μ، σ فإن مركز التوزيع يتغير كما أن تباعد المنحنى حول المركز يتغير أيضا. والشكل (3) يظهر كيف ينتقل مركز المنحنى عندما يتغير المعدل μ من μ_1 ليصبح μ_2. كما يظهر الشكل (4) أن μ ثابتة، لذلك يظل مركز التوزيع ثابتاً ولكن تباعد المنحنى الذي انحرافه المعياري σ_1 أقل من تباعد المنحنى الذي انحرافه المعياري σ_2 لأن $\sigma_1<\sigma_2$ أما الشكل (5) فقد تغير فيه μ, σ.



الشكل (3)



الشكل (4)



الشكل (5)

خواص التوزيع الطبيعي :

1. التوزيع الطبيعي متماثل حول العمود المقام على الوسط μ وشكله يشبه شكل الجرس.

2. للتوزيع الطبيعي قمة واحدة، ولذلك له منوال واحد ينطبق على الوسط μ.

3. يتقارب طرفا منحنى التوزيع الطبيعي من الصفر عندما$x \rightarrow \infty$، $x \rightarrow -\infty$,

4. المساحة تحت منحنى التوزيع الطبيعي تساوي1.

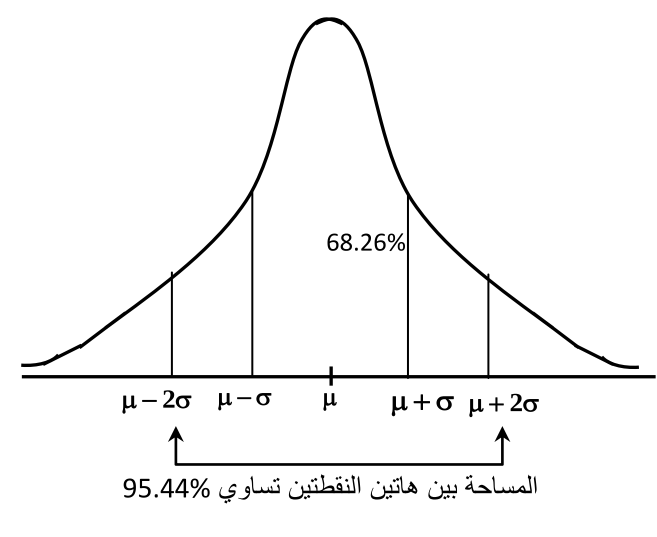
5. هناك نسب معينة من المساحة الواقعة ضمن أي عدد من الانحرافات المعيارية عن الوسط، كما في التوضيحات التالية شكل (6).

المساحة ضمن انحراف معياري واحدة عن الوسط = المساحة الواقعة على الفترة $(\mu - \sigma, \mu + \sigma)$ وتساوي 68.26% من المساحة الكلية. المساحة ضمن $1\frac{1}{2}$ انحراف معياري عن الوسط = المساحة الواقعة على الفترة $(\mu - 1.5\sigma, \mu + 1.5\sigma)$ وتساوي 86.64% من المساحة الكلية.

المساحة ضمن انحرافين معياريين عن الوسط = المساحة الواقعة على الفترة

$(\mu - 2\sigma, \mu + 2\sigma)$ وتساوي 95.44% من المساحة الكلية.

المساحة ضمن ثلاثة انحرافات معيارية عن الوسط = المساحة الواقعة على الفترة $(\mu - 3\sigma, \mu + 3\sigma)$ وتساوي 99.74% من المساحة الكلية.



شكل (6)

6 : 3 التوزيـــع الطبيعـــي المعيـــاري Standard Normal Distribution :

تعريف (2):

التوزيع الطبيعي المعياري هو التوزيع الطبيعي الذي معدله (وسطه) صفر وتباينه 1. فإذا كان المتغير العشوائي Z يخضع للتوزيع الطبيعي المعياري فإن ذلك يعني أن توزيع Z هو التوزيع الطبيعي الذي معدله $\mu = 0$ وتباينه $\sigma^2 = 1$.

نظرية (1) إذا كان التوزيع الاحتمالي للمتغير العشوائي X هو التوزيع الطبيعي ذو المعدل μ والتباين σ^2 فإن توزيع المتغير العشوائي

$$Z = \frac{X-\mu}{\sigma}$$

هو التوزيع الطبيعي المعياري.

وتفيد نظرية (1) أنه إذا كان X ذا توزيع طبيعي بمعدل μ وتباين σ^2 فإن المتغير العشوائي Z الذي نحصل عليه بالتحويل= $Z = \frac{X-\mu}{\sigma}$ يكون ذا توزيع طبيعي معياري (أي ذا توزيع طبيعي معدله صفر وتباينه 1).

كل قيمة من قيم X يقابلها بالطبع قيمة من قيم Z حسب التحويل السابق، وتسمى قيم Z القيم المعيارية المقابلة لقيم X.

مثال (1) :

X متغير عشوائي توزيعه التوزيع الطبيعي ذو المعدل $\mu = 60$ والتباين $\sigma^2 = 49$

حوّل X إلى متغير عشوائي يكون توزيعه التوزيع الطبيعي المعياري.

ثم أوجد القيم المعيارية المقابلة للقيم $x_1 = 65$، $x_2 = 13$، $x_3 = 60$

الحل: المتغير العشوائي الذي توزيعه التوزيع الطبيعي المعياري هو

$$Z = \frac{X-60}{7}$$

القيمة المعيارية المقابلة للقيمة 65 هي: $z_1 = \frac{65-60}{7} = \frac{5}{7}$

القيمة المعيارية المقابلة للقيمة 13 هي

$$z_2 = \frac{13-60}{7} = \frac{-47}{7} = -6.7$$

القيمة المعيارية المقابلة للقيمة 60 هي:

$$z_3 = \frac{60-60}{7} = 0$$

سنعبّر عن العبارة X يخضع لتوزيع طبيعي معدله μ وتباينه σ^2 بالرمز

$X: N(\mu, \sigma^2)$ ولذلك عندما نقول Z يخضع لتوزيع طبيعي معياري فإن ذلك يعني $Z: N(0, 1)$

6 : 4 المساحات تحت التوزيع الطبيعي

بما أن الوسط μ والتباين σ^2 يحدّدان التوزيع الطبيعي فإن المساحة على أي فترة تحت التوزيع الطبيعي (وفوق الخط الأفقي) تعتمد على μ و σ^2. ولما كانت μ، σ^2 تأخذ قيماً كثيرة جداً فإننا لا نستطيع وضع جداول لجميع قيم μ، σ^2.

وبما أنه يمكن تحويل أي توزيع طبيعي إلى توزيع طبيعي معياري فإننا نكتفي باستعمال جدول للتوزيع الطبيعي المعياري.

هناك عدة طرق لعرض جدول التوزيع الطبيعي المعياري منها ما يعطي المساحة إلى يسار القيم المعيارية الموجبة، والسالبة، ومنها ما يعطي المساحة بين $Z = 0$ والقيم الموجبة للمتغير Z. إن جميع هذه الطرق متكافئة ونستعمل هنا الجدول الذي يعطي المساحة إلى يسار قيم Z بشكل عام موجبة كانت أم سالبة.

مثال (1) : إذا كان $Z: N(0, 1)$

أوجد (أ) $P(Z < 1.32)$

(ب) $P(Z > 2.14)$

(جـ) $P(0 < Z < 1.65)$

(د) $P(Z < -1.96)$

(هـ) $P(Z < -2.33)$

$P(Z > -1.3)$ (و)

الحل: باستعمال جدول التوزيع الطبيعي المعياري حيث العمود الأيسر يعطي قيم Z ذات خانة عشرية واحدة والصف العلوي يعطي الخانة العشرية الثانية، نقرأ.

(أ) $P(Z < 1.32) = 0.9066$ من الجدول مباشرة.

(ب) $P(Z > 2.14) = 1 - P(Z \le 2.14) = 1 - 0.9838 = 0.0162$

وذلك باستعمال المتممة واستعمال الجدول

جـ) $P(0 < Z < 1.65) = P(Z < 1.65) - P(Z \le 0)$

وذلك بوضع الاحتمال على شكل الفرق بين مساحتين موجودتين في الجدول مباشرة

$$= 0.9505 - 0.5000 = 0.4505$$

د) $P(Z < -1.96) = 0.025$

هـ) $P(Z < - 2.33) = 0.0099$

و) باستعمال المتممة نجد:

$$P(Z > -1.3) = 1 - P(Z \le -1.3) = 1 - 0.0968 = 0.9032$$

مثال (2) :

إذا كان $Z: N(0, 1)$ أوجد

(أ) $P(-1.7 \le Z \le 2.58)$

(ب) $P(1 \le Z \le 3.3)$

(جـ) $P(-1.23 < Z < -0.68)$

الحل: (أ) نضع الحل كفرق بين مساحتين نجدهما من الجدول مباشرة

$$P(-1.7 \le Z \le 2.58) = P(Z \le 2.58) - P(Z < -1.7)$$

$$= 0.9951 - 0.0446 = 0.9505$$

(ب) $P(1 \le Z \le 3.3) = P(Z \le 3.3) - P(Z < 1)$

$$= 0.9995 - 0.8413 = 0.1582$$

(جـ) $P(-1.23 < Z < -0.68) =$

$$P(Z < -0.68) - P(Z \leq -1.23)$$

$$= 0.2483 - 0.1093 = 0.1390$$

لقد درست كيفية إيجاد الاحتمالات المتعلقة بالمتغير Z ذي التوزيع الطبيعي المعياري والآن إذا كان $X: N(\mu, \sigma^2)$ فكيف تجد الاحتمالات المتعلقة بالمتغير X؟ بالرجوع إلى نظرية (1) تحوّل X إلى Z وتحوّل جميع القيم x إلى قيم معيارية مقابلة لها، وبالتالي تستعمل جدول التوزيع الطبيعي المعياري.

مثال (3): إذا كان $X: N(65 , 36)$ أوجد:

(أ) $P(X > 55)$ (ب) $P(X \leq 68)$

(جـ) $P(50 < X < 70)$ (د) $P(X < 57)$

الحل: حسب النظرية (1) فإن $Z = \frac{X - 65}{6}$ يخضع للتوزيع الطبيعي المعياري، أي أن $Z: N(0, 1)$

(أ) القيمة المعيارية المقابلة للقيمة 55 هي:

$$z = \frac{55 - 65}{6} = -1.67$$

إذن $P(X > 55) = P(Z>-1.67) =1-P(Z \leq -1.67) = 1 - 0.0475 = 0.9525$

(ب) القيمة المعيارية المقابلة للعدد 68 هي $z = \frac{68 - 65}{6} = 0.5$ إذن

$$P(X \leq 68) = P(Z \leq 0.5) = 0.6915$$

مباشرة من الجدول.

(جـ) القيمة المعيارية المقابلة للقيمة 50 هي $z_1 = \frac{50 - 65}{6} = -2.5$

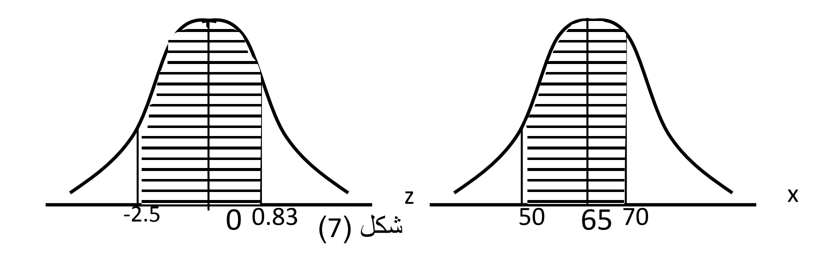
القيمة المعيارية المقابلة للقيمة 70 هي $z_2 = \frac{70 - 65}{6} = 0.83$

إذن

$$P(50 < X < 70) = P(-2.5 < Z < 0.83) = P(Z < 0.83) - P(Z < -2.5)$$

$$= 0.7967 - 0.0062 = 0.7905$$

انظر الشكل (7)



شكل (7)

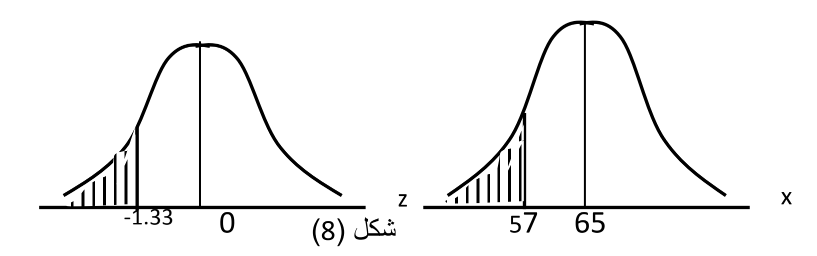
(د) القيمة المعيارية المقابلة للقيمة 57 هي

$$z = \frac{57 - 65}{6} = -1.33$$

$$P(X < 57) = P(Z < -133) = 0.0918$$

مباشرة من الجدول.

انظر الشكل (8)

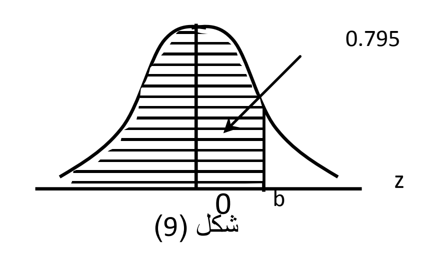


شكل (8)

كما أن جدول التوزيع الطبيعي المعياري يعين المساحات تحت ذلك المنحنى والمحدودة بقيم معينة فإن الجدول نفسه يعطي قيم Z التي تقابلها مساحات معينة إلى يسارها. وذلك بالنظر إلى المساحة المعلومة، وقراءة قيمة Z المطلوبة مع ملاحظة أننا نأخذ أقرب قيمة للمساحة المعطاة.

مثال (4) : إذا كان $Z: N(0, 1)$ أوجد قيمة b بحيث $P(Z \le b) = 0.795$

الحل: بالنظر إلى الشكل المقابل (9) تجد أن المساحة المعطاة هي على الصيغة المعروضة في الجدول ولذلك ننظر مباشرة في الجدول لنبحث عن أقرب قيمة للعدد 0.795 فنجد أنها 0.7939. أقرأ قيمة Z المقابلة فتجدها 0.82 إذاً b = 0.82

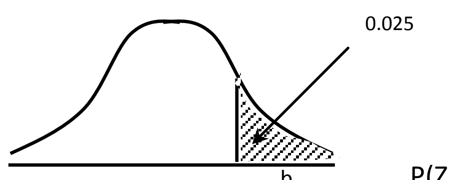


شكل (9)

لاحظ أنه عند البحث عن القيمة القريبة من 0.795 نبحث في داخل الجدول أي في جسم الجدول، وليس في الخط العمودي الأيسر والخط الأفقي العلوي.

مثال (5) : Z:N (0, 1) أوجد b بحيث P(Z > b) = 0.025

الحل: بالنظر إلى الشكل المقابل تجد أن المساحة المعطاة ليست على الصيغة في الجدول ولكنها متممة لها.



الآن P(Z > b) = 0.025

تعطينا P(Z ≤ b) = 0.975 بإيجاد المتممة للطرفين. من الجدول مباشرة نجد

b = 1.96

أما في حالة التوزيع الطبيعي ذي المعدل μ والتباين σ^2 فإنه إذا علمت المساحة وأردت إيجاد قيمة X المقابلة فإن إحدى الطرق السهلة هي أن تحل المسألة في حالة التوزيع الطبيعي المعياري ثم تحوّل القيمة المعيارية التي وجدتها إلى قيمة عاديّة، فتحصل على قيمة X المطلوبة.

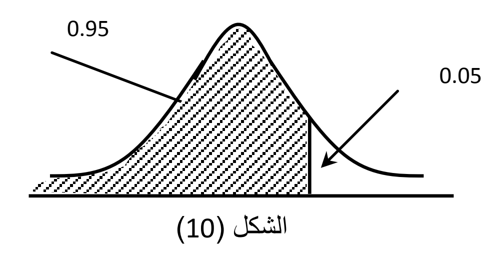
مثال (6) :

إذا كان X:N (70,225)

أوجد a بحيث $P(X > a) = 0.05$

الحل: أوجد أولاً b بحيث $P(Z > b) = 0.05$

من الشكل المقابل (10) يظهر أن المساحة التي على صيغة الجدول هي المساحة المظللة وتساوي 0.95



الشكل (10)

أي أن:

$$P(Z \leq b) = 1 - P(Z > b) = 1 - 0.05 = 0.95$$

بالنظر إلى جسم الجدول الطبيعي المعياري، نجد قيمتين قريبتين من 0.95 هما 0.9495 وتقابلها Z = 1.64، 0.9505 وتقابلها Z = 1.65 ولذلك نأخذ القيمة المتوسطة بينها Z = 1.645 الآن ، 1.645 هي القيمة المعيارية للقيمة a ، أوجد a مــن

المعادلة $\frac{a-70}{15} = 1.645$ أي أن $a = 15 \times 1.645 + 70 = 94.675$ تقريباً.

6 : 5 التوزيع الطبيعي كتقريب لتوزيع ذات الحدين

Normal Distribution as an approximation to the Binomial Distribution

درست الفصل الخامس كيفية حساب الاحتمالات للمتغير العشوائي لذات الحدين من المعادلة

$$b(x; n, p) = P(X = x) = \binom{n}{x} p^x q^{n-x}$$

حيث $q = 1 - p$، p احتمال النجاح في المحاولة الواحدة، n عدد المحاولات.

وعند رسم المدرج الاحتمالي لذات الحدين نلاحظ أن مساحة المستطيل الذي منتصف قاعدته x هي $b(x; n, p)$. فإذا كان $n = 15$ مثلاً، $p = \frac{1}{2}$ فإن مساحة المستطيل الذي منتصف قاعدته 10 يساوي ($b(10; 15, \frac{1}{2}$ أي تساوي

$$\binom{15}{10} (\frac{1}{2})^{10} (\frac{1}{2})^{5}$$

إذا نظرت إلى المدرج الاحتمالي لذات الحدين الذي قيمة n فيه كبيرة الحجم واحتمال النجاح p قريبا من $\frac{1}{2}$ تلاحظ أن بمقدورك تطبيق المنحنى الطبيعي على ذلك المدرج الاحتمالي. ويكون هذا التطبيق جيدا كلما كبرت n وكان المدرج الاحتمالي غير ملتو التواء شديداً ومن الطبيعي أنه إذا كانت p قريبة من $\frac{1}{2}$ فإن المدرج الاحتمالي لا يكون ملتوياً التواء شديداً، وبالتالي: إذا كانت n كبيرة و p قريبة من $\frac{1}{2}$ أمكن تطبيق المنحنى الطبيعي على المدرج الاحتمالي لذات الحدين بشكل جيد. أما التوزيع الطبيعي الذي نستعمله فيجب أن يكون وسطه (معدله) وتباينه هما نفس وسط وتباين ذات الحدين على التوالي، أي أن

$$\mu = np \qquad \sigma^2 = npq$$

ويمثل الشكل (11) المدرج الاحتمالي لمتغير ذات الحدين X الذي فيه $n = 10$ و

$p = 0.5$ وعلى الشكل نفسه تم تطبيق توزيع طبيعي فيه

$$\mu = np = 10 \times \frac{1}{2} = 5$$

$$\sigma^2 = npq = 10 \times \frac{1}{2} \times \frac{1}{2} = 2.5$$

وللمقارنة بين حساب الاحتمالات لمتغير ذات الحدين بدقة من التوزيع الاحتمالي لذات الحدين وحسابها بالتقريب بالتوزيع الطبيعي ندرس المثال التالي:

مثال (7)

في توزيع ذات الحدين $b(x; 10, \frac{1}{2})$

احسب P(X = 6) بدقة ثم بواسطة التقريب بالتوزيع الطبيعي

الحل: $P(X = 6) = \binom{10}{6}\left(\frac{1}{2}\right)^6\left(\frac{1}{2}\right)^4 = \frac{210}{1024} = 0.2005$

وهي مساحة المستطيل الذي منتصف قاعدته 6 أي الذي قاعدته [6.5 - 5.5]. عند تقريب P(X = 6) بواسطة التوزيع الطبيعي، يجب إيجاد المساحة تحت التوزيع. الطبيعي على الفترة [6.5 - 5.5]

أي أن: $P(X = 6) = P(5.5 \le X \le 6.5)$

ولكي تجد هذا الاحتمال، تحول X إلى طبيعي معياري وتجد القيم المعيارية المقابلة للعددين 5.5، 6.5 حيث: X: N (5, 2.5) تقريباً.

لأن $\mu = 10 \times \frac{1}{2} = 5$

$$\sigma^2 = 10 \times \frac{1}{2} \times \frac{1}{2} = 2.5$$

ولذلك فإن: $Z = \frac{X - 5}{1.58}$ يكون توزيعه الطبيعي المعياري

وتكون القيمة المعيارية المقابلة للعدد 5.5 هي $z_1 = \frac{5.5 - 5}{1.58} = 0.32$

القيمة المعيارية المقابلة للعدد 6.5 هي $z_2 = \frac{6.5 - 5}{1.58} = 0.95$

إذن $P(5.5 \le X \le 6.5) = P(0.32 \le Z \le 0.95)$

ومن الجدول

$$= 0.8289 - 0.6255 = 0.2034$$

وكما يظهر فإن الفرق بين الإجابتين أقل من 0.002 حيث = 0.2034 -
0.0016=0.205

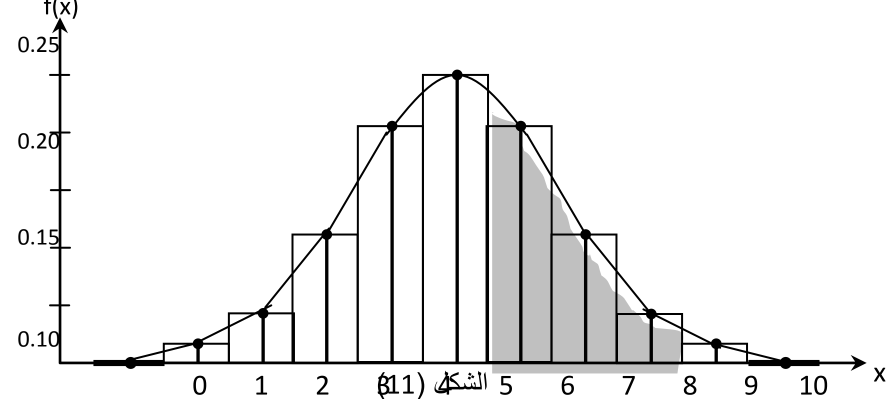
إن عملية طرح 0.5 من الحد الأيسر وإضافة 0.5 إلى الحد الأيمن للفترة المطلوب إيجاد الاحتمال عليها هو ما يسمى عامل التصحيح لتقريب توزيع متغير عشوائي منفصل بواسطة توزيع متغير عشوائي متصل.

مثال (8) :

في التوزيع المعطى في مثال (7) أي ذات الحدين $b(x; 10, \frac{1}{2})$ جد احتمال أن تكون قيم X ما بين 6، 8 متضمنة.

استخرج الجواب بواسطة توزيع ذات الحدين ثم بواسطة التقريب بالتوزيع الطبيعي وقارن الإجابات.

$$P(6 \le X \le 8) =$$
$$= \binom{10}{6}\left(\frac{1}{2}\right)^{6}\left(\frac{1}{2}\right)^{10-6} + \binom{10}{7}\left(\frac{1}{2}\right)^{7}\left(\frac{1}{2}\right)^{10-7} + \binom{10}{8}\left(\frac{1}{2}\right)^{8}\left(\frac{1}{2}\right)^{10-8}$$
$$= \frac{210}{1024} + \frac{120}{1024} + \frac{45}{1024} = 0.366$$



الشكل (11)

وعند التقريب بالتوزيع الطبيعي وبالنظر إلى الشكل (11) نجد أنه يجب أن نبدأ من 5.5 وحتى 8.5 لتغطي قواعد المستطيلات الثلاثة. وهكذا ترى أن $P(6 \le X \le 8)$ يجب تقريبها بالمساحة تحت التوزيع الطبيعي على الفترة 5.5-8.5 أي أنها تقرب بالاحتمال $P(5.5 \le X \le 8.5)$ حيث X هنا يخضع للتوزيع الطبيعي ذي الوسط 5 والتباين 2.5 إذن

$$P(6 \le X \le 8) = P(5.5 \le X \le 8.5)$$

$$= P\left(\frac{5.5-5}{\sqrt{2.5}} \le \frac{X-5}{\sqrt{2.5}} \le \frac{8.5-5}{\sqrt{2.5}}\right)$$

$$= P(0.32 \le Z \le 2.22)$$

$$= 0.9868 - 0.6217 = .3651$$

لاحظ أن الفرق بين الإجابتين يساوي 0.001 تقريباً.

وهذا يعني أن مطابقة المنحنى الطبيعي على المدرج الاحتمالي لذات الحدين كان جيدا. وتزداد الدقة في المطابقة كلما كبرت n وكانت p قريبة من 0.5.

نلخص النتيجة بالنظرية التالية:

نظرية (2) :

إذا كان X متغير ذات الحدين ذا التوزيع $b(x; n, p)$ فإنه يمكن تقريب توزيع X بالتوزيع الطبيعي ذي المعدل $\mu = np$ والتباين $\sigma^2 = np(1 - p)$ عندما تكون n كبيرة، p قريبة من 0.5، وبعبارة أخرى فإن توزيع المتغير $Z = \dfrac{X-np}{\sqrt{np(1-p)}}$ يقرب من التوزيع الطبيعي المعياري $N(0, 1)$

وتقرب احتمالات ذات الحدين باستعمال التوزيع الطبيعي حسب القواعد التالية:

1. إذا كان k عددا صحيحا موجبا بين n, 0 متضمنة فإن:

$$P(X = k) = P(k - \frac{1}{2} \le X \le k + \frac{1}{2})$$

وبالتحويل إلى الطبيعي المعياري يكون:

$$P(X = k) = P\left(\frac{k-\frac{1}{2}-np}{\sqrt{np(1-p)}} \le Z \le \frac{k+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right)$$

حيث $Z: N(0, 1)$

2. إذا كان a، b عددين صحيحين موجبين بين n, 0 متضمنة فإن:

$$P(a \leq X \leq b) = P\left(\frac{a-\frac{1}{2}-np}{\sqrt{np(1-p)}} \leq Z \leq \frac{b+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right)$$

حيث $Z: N(0, 1)$

3. لحساب كل من $P(a < X \leq b), P(a \leq X < b), P(a < X < b)$.

فيجب تحويلها أولاً إلى احتمال فترة مغلقة ثم استعمال (2).

لاحظ أن $P(a \leq X < b)$ تعني احتمال الحصول على X من a متضمنة وحتى b غير متضمنة، فإذا قلت مثلا أوجد $P(7 \leq X < 10)$ فهذا يعني احتمال أن تأخذ X القيمة 7 ثم 8 ثم 9 (فهي لا تأخذ القيمة 10) وبذلك عند التقريب بالتوزيع الطبيعي يكون :

$$P(7 \leq X < 10) = P(7 \leq X \leq 9) = P(6.5 \leq X \leq 9.5)$$

وبعدها نحول X إلى الطبيعي المعياري والقيمتين 6.5 و 9.5 إلى القيم المعيارية.

6 : 6 تطبيقات على التوزيع الطبيعي

Applications on the Normal Distribution

نعطي في هذا البند بعض الأمثلة كتطبيقات على استعمال التوزيع الطبيعي.

مثال (9) :

تخضع أوزان عبوات إحدى أنواع الحلوى لتوزيع طبيعي وسطه 85غم وانحرافه المعياري 2.5غم

أ) ما هو احتمال أن وزن إحدى العبوات الذي أخذته عشوائيا تزيد على 90غم؟

ب) ما هو احتمال أن وزن أحد العبوات الذي أخذته عشوائيا تقل عن 82غم؟

الحل: افرض X = وزن العبوة، فيكون توزيع X التوزيع الطبيعي ذا الوسط 85 والانحراف المعياري 2.5. أي أن $X: N(85, (2.5)^2)$

المطلوب: $P(X > 90)$

نحوّل 90 إلى قيمة معيارية وهي $\frac{90-85}{2.5} = 2$

إذن

$$P(X > 90) = P(Z > 2) = 1 - P(Z \le 2) = 1 - 0.9772 = 0.0228$$

ب.

$$P(X < 82) = P(Z < \frac{82-85}{2.5}) = P(Z < -1.2) = 0.1151$$

مثال (10) :

تخضع تكاليف الولادات الطبيعية في المستشفيات في بلد ما لتوزيع طبيعي وسطه 115 ديناراً وتباينه 49 دينار تربيع.

ما احتمال أن تكون تكاليف إحدى الولادات الطبيعية ما بين 104، 122 ديناراً؟

الحل: افرض تكاليف الولادة تساوي X

إذن $X: N(115, 49)$

المطلوب $P(104 < X < 122)$

بتحويل X إلى طبيعي معياري وتحويل 104، 122 إلى قيم معيارية نجد:

$$P(104 < X < 122) = P\left(\frac{104-115}{7} < \frac{X-115}{7} < \frac{122-115}{7}\right)$$

$$= P(-1.57 < Z < 1) = P(Z < 1) - P(Z \le -1.57)$$

$$= 0.8413 - 0.0582 = 0.7831$$

مثال (11) :

إذا كانت العلامات النهائية للطلبة في أحد المساقات تخضع لتوزيع طبيعي ذي الوسط 68 والانحراف المعياري 12.

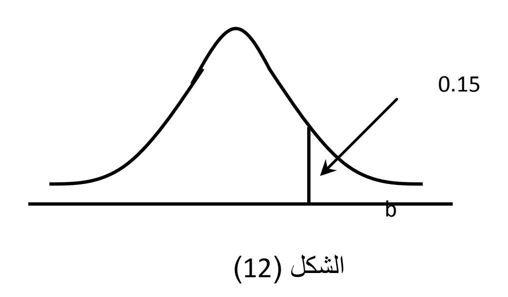
وإذا كان أعلى 15% من الطلبة يحصلون على تقدير "ممتاز"، فما هي أقل علامة تحصل على تقدير "ممتاز"؟

الحل: افرض X هي علامة أي طالب في ذلك المساق,

إذن: $X: N(68, 12^2)$

المطلوب إيجاد b بحيث $P(X \geq b) = 0.15$

انظر الشكل (12).



الشكل (12)

نجد أولا a بحيث $P(Z \geq a) = 0.15$ حيث Z يخضع للتوزيع الطبيعي المعياري.

نأخذ المتممة للطرفين فتحصل على

$P(sZ < a) = 0.85$

إذن $a = 1.04$ تقريباً، وهذه هي القيمة المعيارية للعدد b

إذن $1.04 = \dfrac{b - 68}{12}$

$$b = 68 + 12 \times 1.04 = 80.48$$

وهي أقل علامة تحصل على "ممتاز"؟.

6 : 7 التحقق من التوزيع الطبيعي Normal Proability Plots

لقد درسنا في البنود السابقة كيف نتعامل مع التوزيع الطبيعي مثل حساب الاحتمالات أو النسب المئوية واستعمال الجداول وإيجاد المئينات.

هناك اهتمامات أخرى يحتاج إليها الباحث الإحصائي مثل تقرير فيما إذا كان مجتمع ما خاضعاً لتوزيع طبيعي أم لا، أو فيما إذا كان توزيع أحد المتغيرات العشوائية هو التوزيع الطبيعي أم لا.

سندرس في الفصول القادمة، في الإحصاء الاستدلالي، نظريات عديدة تفترض أن المجتمع يخضع لتوزيع طبيعي، ولذلك بات من الضروري معرفة كيف نتحقق أن توزيع المجتمع تحت الدراسة هو توزيع طبيعي أم لا.

إن دراسة أشكال التوزيعات التكرارية للبيانات التي تتضح من المدرج التكراري أو المضلع التكراري أو العرض بطريقة الغصن والورقة تظهر لنا فيما إذا كان توزيع البيانات قريباً من التوزيع الطبيعي أم لا. أي أننا نستطيع مقارنة شكل التوزيع التكراري مع شكل التوزيع الطبيعي من حيث كونه كشكل الجرس، ذي منوال واحد، متماثل حول الوسط. فمن المعلوم أن العينة المأخوذة من توزيع طبيعي سيكون توزيعها التكراري قريباً من التوزيع الطبيعي وخاصة إذا كان حجم العينة كبيراً، أما إذا كان حجم العينة صغيراً فإن التوزيع التكراري أو شكل الغصن والورقة لهذه العينة لا يعكس بالضرورة شكلاً يشبه منحنى التوزيع الطبيعي، ولذلك نحتاج إلى طرق أكثر دقة لنتعرف فيما إذا كانت العينة قد أخذت من توزيع طبيعي. إن رسم الاحتمالات الطبيعية تعطينا إحدى هذه الطرق الدقيقة.

تتلخص طريقة رسم الاحتمالات الطبيعية برصد البيانات من العينة المعطاة مقابل البيانات التي نتوقع الحصول عليها لو كانت العينة ذات الحجم نفسه قد أخذت من توزيع طبيعي معياري وتسمى بيانات هذه العينة القيم الطبيعية المعيارية Normal Scores.

إذا كانت العينة تحت الدراسة من مجتمع طبيعي فإن رسم الاحتمالات الطبيعية سيكون خطياً تقريباً (يكون الرسم خطاً مستقيماً تقريباً) والعكس بالعكس.

ولدى تطبيق هذه الطريقة علينا أن نتذكر أمرين، أولهما أن الحكم فيما إذا كان الرسم خطاً مستقيماً تقريباً أمر شخصي، وثانيهما يجب الأخذ بالاعتبار أننا نبني الحكم بناء على عينة من المجتمع.

ومع أخذ هذه الأمور بعين الاعتبار فإن تقرير أن عينة ما قد أخذت من مجتمع طبيعي باستعمال رسم الاحتمالات الطبيعية يكون تحت الإرشادات التالية:

للحكم بأن عينة قد أخذت من مجتمع طبيعي، نبني رسم الاحتمالات الطبيعية للعينة ونلاحظ

(1) إذا كان الرسم خطياً تقريباً نقول إن المجتمع طبيعي تقريباً.

(2) إذا كان الرسم غير خطي أي يظهر انحناء واضحاً، نقول ربما أن المجتمع لا يخضع لتوزيع طبيعي.

إذا كان حجم العينة صغيراً وجب النظر لهذه الإرشادات بشكل فضفاض أما إذا كان حجم العينة كبيراً أمكن تطبيق الإرشادات بشكل دقيق.

مثال (12) :

تمثل العينة التالية مداخيل 12 موظفاً بآلاف الدنانير في سنة 1998.

16.5	66.4	18.1
28.0	45.5	10.6
70.0	9.8	12.5
13.5	40.0	50.8

أ) أرسم الاحتمالات الطبيعية لهذه البيانات.

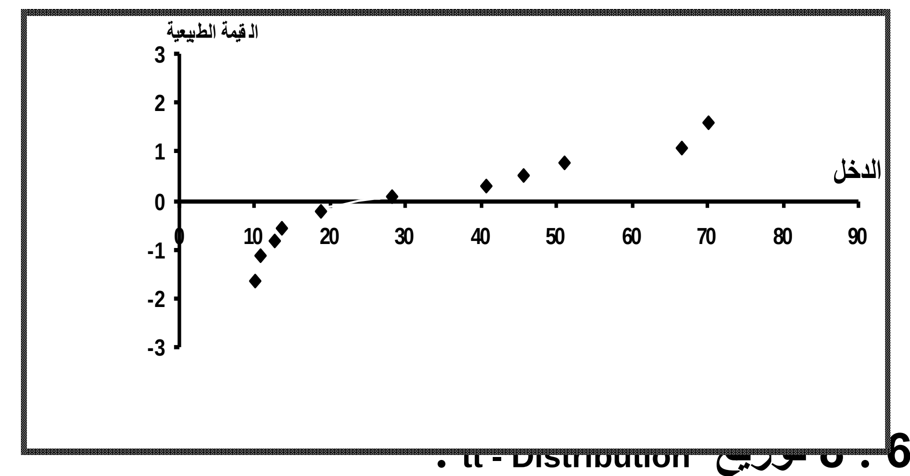
ب) هل هذه العينة مأخوذة من توزيع طبيعي (أي هل مداخيل الموظفين السنوية تخضع لتوزيع طبيعي؟)

الحل: نرتب البيانات تصاعدياً ونرتّب مقابلها القيم الطبيعية المعيارية المأخوذة من الجدول الخاص بذلك المقابل للحجم n = 12.

القيم الطبيعية المعيارية	المداخيل
-1.64	9.8
-1.11	10.6
-0.79	12.5
-0.53	13.5
-0.31	16.5
-0.10	18.5
0.10	28.0
0.31	40.5
0.53	45.5
0.79	50.8
1.11	66.4

70.0	1.64

نرصد النقاط التي احداثياتها الأفقية المداخيل واحداثياتها العمودية القيم الطبيعية أي نرصد النقاط:(9.8,-1.64),(10.6,-1.11),...,(66.4,1.11),(70.0,1.64) ونحكم فيما إذا كانت هذه النقاط تمثل خطاً مستقيماً تقريباً. نلاحظ أن الرسم البياني يظهر انحناء واضحاً. أنظر الشكل (13).



6 . 8 توزيع t - Distribution .

إن أحد التوزيعات المتصلة الهامة هو توزيع t، وهو توزيع لمتغير عشوائي متصل.

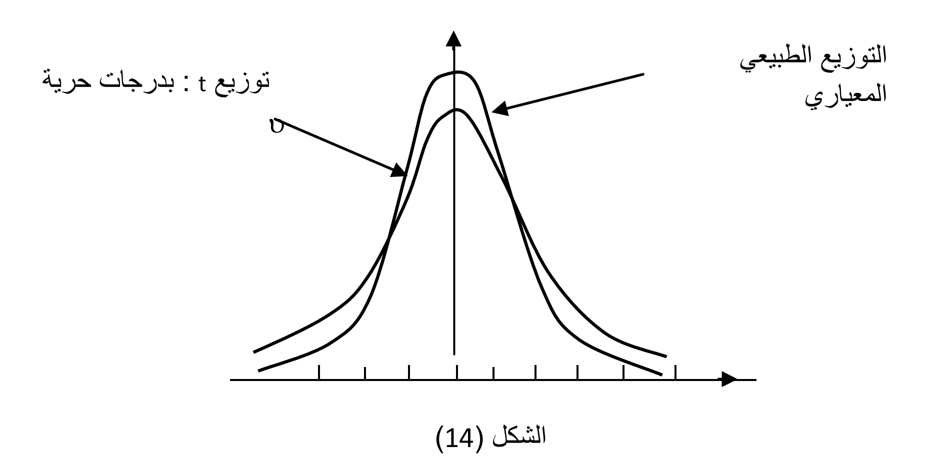
تعريف (3) : إذا كان توزيع الكثافة الاحتمالي للمتغير العشوائي t معطى بالمعادلة :

$$\mathbf{f(t) = c\left(1+\frac{t^2}{\upsilon}\right)^{-\upsilon+\frac{1}{2}}} \quad , \quad -\infty < t < \infty$$

فإن هذا التوزيع يسمى توزيع t حيث υ درجات الحرية و c ثابت يعتمد على υ ليجعل المساحة تحت المنحنى تساوي 1.

يشبه منحنى توزيع t شكل الجرس، وهو أحادي المنوال له قيمة تقابل t = 0 وهو متماثل حول العمود المقام على t = 0.

ويشبه شكله شكل التوزيع الطبيعي المعياري إلا أنه أكثر انخفاضا منه. هذا إضافة إلى أن تقارب طرفيه من الصفر عندما $t \to \infty$, $t \to -\infty$ أبطأ من تقارب طرفي التوزيع الطبيعي المعياري. ويعتمد منحنى توزيع t على معلمة هامة تحدد شكل ذلك المنحنى ، وهي درجات الحرية. فعندما يزداد عدد درجات الحرية يقرب توزيع t من التوزيع الطبيعي المعياري. في الشكل (14) يظهر كل من منحنى توزيع t ذي n درجات حرية، ومنحنى التوزيع الطبيعي المعياري وذلك لتسهيل المقارنة بينهما.



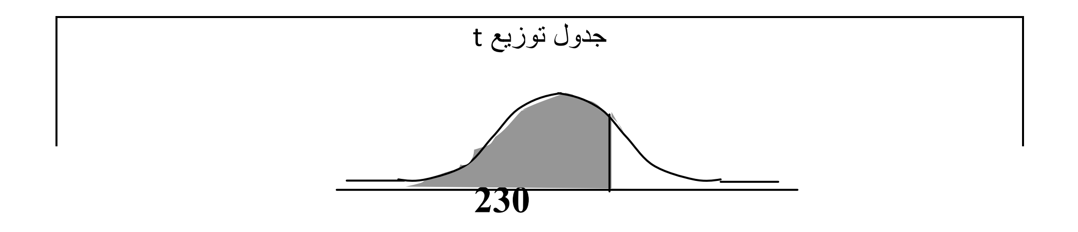
الشكل (14)

نحسب الاحتمالات تحت توزيع t وذلك بحساب المساحات تحت منحنى ذلك التوزيع مع معرفة درجات الحرية له. وهناك جداول خاصة لهذه المساحات حيث تسجل درجات الحرية في العمود الأيسر، وعلى الخط الأفقي تسجل مساحات معينة. أما الأعداد داخل الجدول فتمثل قيم t المقابلة لدرجات حرية معينة والتي تقع المساحة المعينة إلى يسارها.

فمثلاً، إذا نظرت إلى القيمة 0.90 على الخط الأفقي العلوي، والعدد 5 في العمود الأيسر الذي يمثل درجات الحرية، فإنك تجد أن العمود النازل من القيمة 0.90 يتقاطع من الخط الأفقي المار من 5 عند القيمة 1.476 وهذا يعني أن 1.476 هي قيمة t التي يقع إلى يسارها 0.90 من المساحة تحت توزيع t ذي 5 درجات حرية.

أنظر الجدول التالي الذي يمثل جزءا من جدول توزيع t.

جدول توزيع t



$t[\lambda;\upsilon]$

المساحة المظللة تمثل $P(-\infty < t < t[\lambda;\upsilon])$

حيث υ هي درجات الحرية لتوزيع t .

λ / υ	0.75	0.90	0.95	0.975	0.99	0.995	0.9975	0.999
1	1.000	3.078	6.314	12.706	31.821	63.657	127.321	318.309
2	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.327
3	0.765	1.638	2.353	3.186	4.541	5.841	7.453	10.214
4	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173
5	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893
6	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208
7	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785

ولو نظرت إلى الجدول على القيمة 0.95 في السطر الأفقي العلوي، ونظرت إلى درجات الحرية 7 في العمود الأيسر ، لوجدت أن العمود النازل من 0.95 بتقاطع مع الخط الأفقي المار من 7 عند القيمة 1.895 وهذا يعني أن t = 1.895 هي قيمة t ذي درجات حرية 7 التي يقع إلى يسارها 0.95 من المساحة.

أما إذا أردت إيجاد المساحة الواقعة إلى يسار قيمة t تحت توزيع t ذي درجات حرية معينة، فعليك أن تنظر على درجات الحرية المعينة وتسير أفقيا إلى أن تصل العدد الأقرب للقيمة المعطاة لديك ، ثم تصعد من هذه القيمة إلى أعلى لتصل إلى الخط الأفقي العلوي. والقيمة التي تصل إليها على الخط الأفقي العلوي هي المساحة إلى يسار قيمة t المعطاة.

فمثلا لإيجاد المساحة إلى يسار 4.541 تحت توزيع t ذي 3 درجات حرية فإننا نسير أفقيا بادئين من درجات الحرية 3 إلى أن نصل العدد 4.541 فننظر رأسيا فنجد المساحة 0.99.

نعبر عن قيمة t التي يكون إلى يسارها مساحة معينة قيمتها λ تحت منحنى توزيع t ذي درجات حرية m بالرمز $t[\lambda ; m]$ كما يظهر في الشكل وعند النظر إلى الجدول في الملحق، تجد أنه يعطي قيم $t[\lambda ; m]$ لقيم λ القريبة من 1، أما عندما تكون λ صغيرة مثل 0.01 , 0.05 وغيرها، نستعمل القاعدة $t[\lambda ; m] = -t[1-\lambda ; m]$ ، وذلك بسبب تماثل توزيع t ، حول العمود المقام على الصفر.

مثال (13) :

المتغير العشوائي t، يخضع لتوزيع t، ذي درجات الحرية 4، أوجد :

أ- المساحة الواقعة على يسار 1.532.

الحل : بالنظر إلى درجات الحرية 4 في جدول توزيع t، والسير أفقيا تجد أن أقرب عدد للقيمة المعطاة هو 1.533 وبالصعود عموديا نلاحظ أن المساحة 0.90.

ب- ما هي قيمة t التي يقع إلى يسارها مساحة 0.01.

الحل : بالنظر إلى 4 في العمود الأيسر و 0.99 في الخط الأفقي نجد القيمة 3.747 وهي القيمة التي يقع إلى يسارها مساحة 0.99 وبماأنك تريد القيمة التي يقع إلى يسارها مساحة 0.01 فإن القيمة تساوي 3.747- بالتماثل.

(جـ) أوجد λ بحيث

$$t\,[\lambda : 4] = -2.776$$

الحل : بالتماثل $t\,[\lambda ; 4] = -t[1-\lambda ; 4]$ ولكن $t[1-\lambda ; 4] = 2.776$

ومن الجدول مباشرة : $1-\lambda = 0.975$ إذن $\lambda = 0.025$

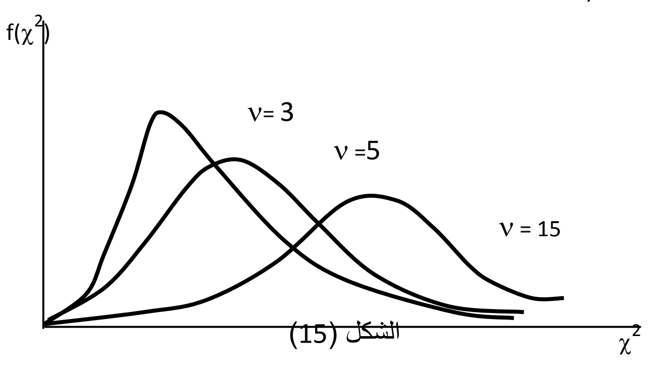
6 : 9 توزيع كاي تربيع (Chi-Square Distribution) :

تعريف (4) : إذا كان توزيع الكثافة الاحتمالي للمتغير العشوائي χ^2 معطى بالمعادلة :

$$\mathbf{f}(\chi^2) = \mathbf{c}(\chi^2)^{(\nu-2)/2,}\mathbf{e}^{-\chi^2/2},\ \chi^2 > 0$$

فإن هذا التوزيع يسمى توزيع كاي تربيع ذا درجات حرية ν حيث c تعتمد على ν وتحدد بحيث تكون المساحة تحت المنحنى 1.

إن المعادلة السابقة تعين لنا منحنى توزيع كاي تربيع والشكل (16) يمثل منحنى ذلك التوزيع على درجات $\nu = 3$, $\nu = 5$, $\nu = 15$.

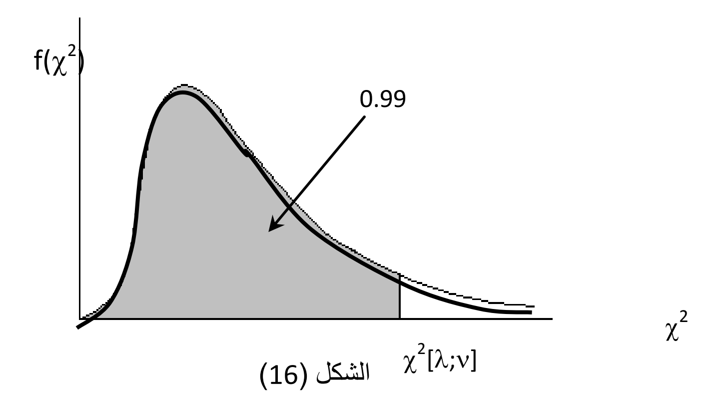


الشكل (15)

ولإيجاد المساحات تحت منحنى كاي تربيع أو إيجاد القيم التي يقع إلى يسارها أو إلى يمينها مساحة معينة نستعمل جدول كاي تربيع حيث يسجل عدد درجات الحرية في العمود الأيسر وتسجل المساحات إلى يسار قيمة χ^2 على الخط الأفقي وتسجل قيم χ^2 داخل الجدول.

نعبر عن قيمة χ^2 التي يقع إلى يسارها مساحة λ تحت منحنى توزيع χ^2 على درجات حرية ν بالرمز $\chi^2\,[\lambda\,;\,\nu]$ باستعمال توزيع كاي تربيع على درجات حرية 10 أوجد :

أ- قيمة χ^2 التي يكون إلى يسارها 0.99 من المساحة.



الشكل (16)

الحل : بالنظر إلى درجات حرية 10 والنظر إلى المساحة 0.99 على الخط الأفقي نجد نقطة التقاطع 23.209 وهي القيمة المطلوبة ، أي أن :

$\chi^2\,[0.99;\,10] = 23.209$ أنظر الشكل (16).

ب- قيمة χ^2 التي يكون إلى يمينها 0.01 من المساحة.

الحل : النقطة التي يكون إلى يمينها 0.01 من المساحة هي النقطة التي يكون إلى يسارها 0.99 من المساحة، فالقيمة المطلوبة هي :

$$\chi^2\,[0.99;\,10] = 23.209$$

جـ- قيمة χ^2 التي يكون إلى يسارها 0.975 والقيمة التي يكون إلى يسارها 0.025 من المساحة :

$$\chi^2\,[0.975\,;\,10] = 24.483$$

$$\chi^2\,[0.025\,;\,10] = 3.247$$

6 : 10 توزيع F : (The F Distribution) :

من التوزيعات الاحتمالية الهامة التي تستعمل في اختبار الفرضيات هو توزيع F.

تعريف (5) : إذا كان توزيع الكثافة الاحتمالي للمتغير العشوائي F معطى بالمعادلة :

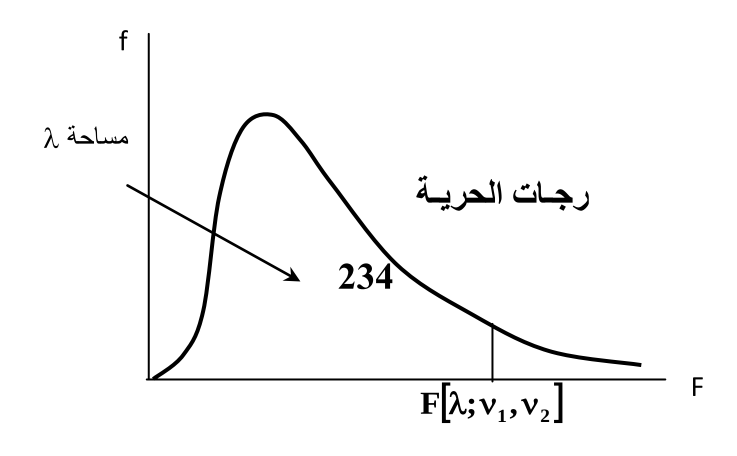
$$\mathbf{f(F)} = \frac{\mathbf{cF}^{(\nu_1-2)/2}}{(\nu_2+\nu_1\mathbf{F})^{(\nu_1+\nu_2)/2}}, \mathbf{F} > 0$$

فإن هذا التوزيع يسمى توزيع F ويعبر عنه بالرمز $F(\nu_1, \nu_2)$ حيث ν_1 وν_2 هي درجات الحرية و c ثابت يعتمد على ν_1 وν_2 ويعين بحيث تصبح المساحة تحت منحنى التوزيع تساوي 1.

يوجد لهذا التوزيع عددان من درجات الحرية، وبما أن ν_2 يظهر في المقام فقط فإنه يعتبر درجات حرية المقام ويعتبر ν_1 درجات حرية البسط.

إن اقتران الكثافة الاحتمالي يعين الرسم البياني لمنحنى توزيع F الذي يعتمد على درجات الحرية ν_1 و ν_2.

الشكل(17) : يعطي منحنى هذا التوزيع ونلاحظ انه عندما يكون $\nu_1 > 2$ و $\nu_2 > 2$ فإن توزيع F أحادي المنوال ملتو إلى اليمين قليلا. وكلما ازدادت درجات الحرية ν_1 و ν_2 يقرب توزيع F من التوزيع الطبيعي، وهو موجب لجميع قيم F بين الصفر واللانهاية.



الشكل (17)

هذا ويمكن استعمال الجدول 5 في الملحق لإيجاد المساحات تحت منحنى توزيع F.

ونستعمل الرمز $F[\lambda;\nu_1,\nu_2]$ ليدل على النقطة على المحور الأفقي التي يكون إلى يسارها مساحة λ كما يظهر في الشكل (17) فمثلا :

$$F[0.95 ; 9 , 7] = 3.68$$

$$F[0.025 ; 11 , 10] = 0.283$$

وإذا عرفت درجات الحرية وقيمة F نستطيع إيجاد المساحة إلى يسار F وإلى يمينها من الجداول أيضا، فمثلا لإيجاد المساحة إلى يسار F = 3 إذا كانت $\nu_1 = 7$ و $\nu_2 = 20$. نلاحظ أن أقرب عدد في الجدول للقيمة 3 هو 3.01 فنجد F [0.975 ; 7 , 20] = 3.01 وتكون المساحة المطلوبة 0.975.

عند قراءتنا لجداول F نلاحظ أن بعض القيم غير موضوعة فيها مثل $F[0.05;\nu_1,\nu_2]$ أو $F[0.01;\nu_1,\nu_2]$، ولإيجاد هذه القيم نستعمل القاعدة :

$$F[\lambda;\nu_1,\nu_2] = \frac{1}{F[1-\lambda;\nu_2,\nu_1]}$$

لاحظ أن المقام هو قيمة المتممة للمساحة المطلوبة مع تبديل درجات الحرية فيما بينها.

مثال (15) : أوجد : أ- $F[0.05; 10 ,7]$ ، ب- $F[0.01; 5 ,10]$.

الحل :

(أ)
$$F[0.05; 10 ,7] = \frac{1}{F[0.95; 7, 10]}$$

$$= \frac{1}{3.14} = 0.318$$

(ب) $$F[0.01; 5, 10] = \frac{1}{F[0.99; 10, 5]}$$

$$= \frac{1}{10.1} = 0.099$$

تمارين

1-6 : إذا كان $Z: N(0, 1)$

أوجد

$P(Z < 1.29)$

$P(Z > 2.56)$

$P(-1.1 \leq Z \leq 3.2)$

$P(2.7 < Z < 4.3)$

$P(-1.96 \leq Z \leq 1.96)$

$P(Z > 1.64)$

$P(Z \leq -1.65)$

$P(Z \leq -2.33)$

$P(Z \geq 2.58)$

$P(-1.28 \leq X \leq 1.28)$

2-6: إذا كان $Z: N(0, 1)$

أوجد a في كل مما يلي:

$P(Z \leq a) = 0.99$

$P(Z > a) = 0.01$

$P(-a \leq Z \leq a) = 0.95$

$P(-a \leq Z \leq a)\ 0.99$

$P(Z \leq a) = 0.01$

$P(0 \leq Z \leq a) = 0.49$

$P(a \leq Z \leq 0) = 0.45$

3-6 : إذا كان $X: N(45, 64)$

أوجد

$P(X \le 45)$

$P(X \le 22)$

$P(X \ge 64)$

$P(30 \le X \le 70)$

$P(X \ge 61)$

$P(X \le 10)$

$P(X \ge 80)$

4-6 : إذا كان $X: N(50, 36)$

أوجد b في كل مما يلي

$P(X \le b) = 0.99$

$P(X > b) = 0.05$

$P(X \le b) = 0.01$

$P(X \ge b) = 0.95$

$P(50 \le X \le b) = 0.3$

$P(b \le X \le 50) = 0.15$

5-6 : قرر مدرس الإحصاء أن يعطي أعلى 10% من الطلبة في تلك المادة تقدير "ممتاز". إذا كانت علامات الطلبة تخضع لتوزيع طبيعي وسطه 70 وانحرافه المعياري 15

(أ) فما هي أقل علامة تحصل على تقدير "ممتاز"؟

(ب) إذا كانت علامة الرسوب في تلك المادة دون الخمسين، فما نسبة عدد الطلبة الراسبين؟

(جـ) ما عدد الطلبة الحاصلين على تقدير "ممتاز" إذا كان عدد الطلبة في المادة 300 طالباً.

6-6 : إذا كانت قوة نوع من خيوط سنارة صيد الأسماك تخضع لتوزيع طبيعي وسطه 13كغم وانحرافه المعياري 2كغم.

أ. اخترت خيطاً بطريقة عشوائية، ما احتمال أن تكون قوته أكبر من 16كغم؟

ب. ما احتمال أن تكون قوة الخيط الذي اخترته أقل من 12كغم؟

جـ. ما هو المئين الثمانون لقوة هذه الخيوط؟

7-6 : إذا كانت أطوال الحياة لمصابيح كهربائية ينتجها مصنع معين 800 ساعة وانحرافها المعياري 60 ساعة فإذا كانت أطوال الحياة موزعة حسب التوزيع الطبيعي، فما نسبة عدد المصابيح من إنتاج ذلك المصنع التي تقل أطوال حياتها عن 770 ساعة؟ ما نسبة عدد المصابيح التي تزيد أطوال حياتها عن 900 ساعة؟

8-6 : تخضع معاملات الذكاء للطلبة الجامعيين لتوزيع طبيعي معدله 105 وانحرافه المعياري 10. ما نسبة عدد الطلبة الذين يزيد معامل الذكاء لديهم على 125؟

9-6 : إذا كانت أطوال الجنود في أحد الجيوش موزعة حسب التوزيع الطبيعي الذي وسطه 170سم وانحرافه المعياري 5سم.

أ) ما احتمال أن يزيد طول جندي اخترته عشوائيا على 176سم؟

ب) ما نسبة عدد الجنود الذين تزيد أطوالهم على 173سم؟

جـ) ما نسبة عدد الجنود الذين تقل أطوالهم عن 162سم؟

10-6 : يحتوي امتحان عام على 100 سؤال من نوع "الاختيار من متعدد" ولكل سؤال أربع إجابات، واحدة منها صحيحة. تقدم أحد الطلبة للامتحان وأجاب عن جميع الأسئلة بالتخمين.

أ) ما احتمال أن يحصل الطالب على 50 إجابة صحيحة؟

ضع الإجابة على شكل معادلة ثم أوجد تقريباً لإجابتك باستعمال التوزيع الطبيعي.

ب) ما احتمال حصول الطالب على إجابات صحيحة عددها أكثر من 35 وأقل من 76؟

جـ) ما احتمال حصول الطالب على إجابات صحيحة عددها ما بين 27، 62 متضمّنة؟

6-11 : تخضع كمية القار في نوع من السجاير لتوزيع طبيعي وسطه 18.5ملغم وانحرافه المعياري 2.2ملغم.

أ) ما احتمال أن كمية القار في سيجارة أخذتها عشوائيا تزيد على 16.7ملغم؟

ب) ما هو احتمال أن كمية القار في سيجارة أخذتها عشوائيا تقل عن 15.9ملغم؟

6-12 : تخضع أوزان العلب المعبأة في مصنع للتمور لتوزيع طبيعي وسطه 250غم وانحرافه المعياري 20 غراماً.

اشتريت علبة من منتجات هذا المصنع. ما احتمال أن يقع وزنها ما بين 240، 265غم؟

6-13 : أخذت عينة عشوائية حجمها 12 فأعطت القيم 8.5، 13.5، 8.4،

7.7، 12.3، 9.3، 9.9، 6.3، 9.7، 9.5، 7.0، 6.9

هل كانت هذه العينة من مجتمع طبيعي؟

6-14 : أخذت عينة عشوائية حجمها 4 فأعطت القيم 18.5، 23.5، 16.9، 17.7 هل تدل هذه العينة على أنها من توزيع طبيعي.

6-15 : أوجد النقاط :

t [0.99 ; 10]

t [0.95 ; 15]

t [0.01 ; 10]

t [0.05 ; 15]

t [0.99 ; 8]

t [0.01 ;120]

6-16 : أوجد المئين العاشر P_{10} والمئين العشرين لتوزيع t ذي درجات الحرية 20.

6-17 : أوجد الربيع Q_1 والوسيط والربيع الثالث Q_3 للتوزيع في تمرين 16.

6-18 : أوجد قيمة λ في كل مما يأتي :

$t[\lambda ; 20] = 1.325$

$t[\lambda ; 5] = 2.015$

$t[\lambda ; 12] = 2.681$

$t[\lambda ; 15] = 2.131$

6-19 : أوجد النقاط التالية مع التوضيح بالرسم :

أ- $\chi^2 [0.05 ; 6]$, $\chi^2 [0.95 ; 6]$

ب- $\chi^2 [0.01 ; 15]$, $\chi^2 [0.99 ; 15]$

جـ- $\chi^2 [0.05 ; 20]$, $\chi^2 [0.95 ; 20]$

6-20 : أوجد λ في كل مما يلي :

$\chi^2 [\lambda ; 5] = 11.07$

$\chi^2 [\lambda ; 5] = 1.145$

$\chi^2 [\lambda ; 10] = 2.156$

$\chi^2 [\lambda ; 10] = 4.865$

6-21 : أوجد النقاط التالية مع التوضيح بالرسم :

$F [0.95 ; 7, 10]$

$F [0.05 ; 10, 7]$

$F [0.99 ; 8 , 5]$

$F [0.01 ; 5 , 8]$

6-22 : أوجد قيمة λ في كل مما يأتي مع التوضيح بالرسم :

$F [\lambda ; 8 ; 12] = 4.50$

$F [\lambda ; 2 ; 24] = 0.025$

$F[\lambda ; 24 ; 3] = 0.269$

$F[\lambda ; 9 ; 11] = 4.63$

$F[\lambda ; 8 ; 12] = 8.94$

23-6 : في توزيع χ^2 ذي درجات حرية 15 أوجد :

$P(\chi^2 < 5.229)$

$P(\chi^2 > 7.261)$

$P(6.262 < \chi^2 < 27.488)$

24-6 : في توزيع t على درجات حرية 12 أوجد :

$P(t > 2.681)$

$P(t < 3.055)$

$P(-2.179 < t < 2.179)$

$P(-1.782 < t < 1.782)$

25-6 : في توزيع χ^2 بدرجات حرية 20 أوجد المئين العاشر والمئين التسعين.

الفصل السابع
العينات وتوزيعات المعاينة
Samples and Sampling Distributions

7 : 1 مقدمة :

من أهم أغراض الإحصاء الاستدلال الذي يمكِّن الباحث من التوصل إلى التعميمات عن طريق دراسة المشاهدات في العينة. ترتكز هذه التعميمات على فهم السّمات والتغيرات في المجتمع بنقلها عن طريق المعاينة. إلى تغير

في الإحصاءات، مثل الوسط الحسابي والتباين. ومن وجهة عامة، عادة ما يكون اهتمامنا منصبا على معرفة سمات المجتمع (المعلمات) عن طريق إعطائها قيما عددية وذلك باستعمال سمات مشابهة لها خاصة بالعينة (الإحصاءات).

وإمكانية التعميم من العينة إلى المجتمع تعتمد على كيفية أخذ العينة وحجمها وطرق دراسة صفاتها باستعمال نظرية الاحتمال.

تقودنا هذه المقدمة إلى دراسة عدد من المفاهيم ومنها المجتمع، العينة ، طرق أخذ العينات، المعلمات، الإحصاءات ..... الخ.

تعريف (1)

المجتمع هو مجموعة العناصر أو الأفراد الذي ينصب عليهم الاهتمام في دراسة معينة، أو مجموعة المشاهدات أو القياسات التي تم جمعها عن تلك العناصر.

تعريف (2)

وحدة المعاينة: هي أي عنصر أو فرد في المجتمع الذي ندرسه.

تعريف (3)

العينة: هي مجموعة جزئية من المجتمع.

7 : 2 طرق جمع البيانات الإحصائية

يتم جمع البيانات الإحصائية بإحدى الطرق التالية:

(1) طريقة المسح الشامل Census

فيها تجمع البيانات من جميع أفراد المجتمع الإحصائي، فمثلا:

إذا أردنا القيام بتعداد عام للسكان في بلد قليل السكان نسبياً، كالأردن مثلا، فإننا نقوم بعد أفراد جميع العائلات في المجتمع الأردني، وإذا أردنا التعرف على مستوى طلاب جامعة اليوموك في الرياضيات فإننا نقوم برصد علامات جميع هؤلاء الطلاب في الرياضيات لدى المسجل العام، أو نقوم بتصميم فحص عام في الرياضيات نطبقه على جميع طلبة الجامعة ومن خلال نتائجهم في هذا الفحص، نحاول التعرف على مستوى طلبة الجامعة في الرياضيات.

(2) طريقة العينة Sample

هناك عدة حالات يتعذر فيها المسح الشامل وعندها نلجأ إلى دراسة جزء من المجتمع الإحصائي يسمى "العينة". وحجم العينة هو عدد عناصرها.

ومن الأمثلة على الحالات التي يصعب فيها إجراء المسح الشامل أو يتعذر:

أ) فساد عناصر المجتمع نتيجة أخذ المشاهدات من تلك العناصر، فمثلا:

إذا أردنا معرفة مدى صلاحية البيض الذي تنتجه مزرعة ما للأكل فمن غير المعقول أن نقوم بتكسير كل إنتاج تلك المزرعة من البيض للتأكد من صلاحيته، لذا نجد أن أخذ جزء من هذا البيض لفحصه أمر ضروري ومن ثم نعمم نتيجة الفحص على المجتمع الإحصائي، وهو هنا، إنتاج تلك المزرعة من البيض.

ب) غالبا ما يتعذر الوصول إلى جميع عناصر المجتمع الإحصائي وإجراء الدراسة عليها، فمثلاً: إذا أردنا دراسة كميات الأمطار التي سقطت في الأردن منذ عام 1940 حتى الآن من أجل التعرف على أثرها على كميات إنتاج القمح في الأردن، فقد لا يكون لدينا سجلات كافية تحتوي على المعلومات المطلوبة منذ عام 1940، لذا تكتفي بدراسة عينة من السجلات المتوفرة.

ج) إن دراسة العينة أدنى كلفة من المسح الشامل وتستغرق وقتا وجهدا أقل.

د) قد يتطلب إجراء المسح الشامل عددا كبيرا من الأشخاص المدربين لإتمام جمع البيانات، أما العينة فتحتاج إلى عدد أقل من هؤلاء وبذلك يتم التوفير في عدد العاملين. هذا من جهة، ومن جهة ثانية، إذا كان عدد المطلوبين لإجراء التعداد قليلا فإنه يمكن تدريبهم على جمع البيانات بصورة أفضل، وبالتالي يمكن التقليل من الأخطاء والفروق الفردية في جمع البيانات.

هـ) قد يحتاج الباحث إلى نتائج "سريعة" تعطيه مؤشرا عن ظاهرة معينة، وبالتالي فهو لا يستطيع دراسة المجتمع كله في ذلك الوقت القصير، لذا يلجأ إلى دراسة عينة فقط من ذلك المجتمع. ومن أهم الأمثلة على ذلك استطلاع الرأي العام بعد حوادث معينة، أو قبيل الانتخابات، أو قبيل إجراء استفتاء عام حول موضوع مهم.

و) قد يكون المجتمع متصلا، كأن تكون مجموعة عناصره غير قابلة للعد، وبالتالي يتعذر إجراء مسح شامل لها، مثل: دراسة مخزون الأردن من الفوسفات، أو دراسة مخزون دولة الكويت من البترول، حيث يتعذر إجراء مسح شامل، ويكتفي بدراسة عينة.

وفي بعض الأحيان، وحتى لو كان المجتمع منفصلا، أي أن مجموعة عناصره قابلة للعدّ، فقد لا يكون من السهل إجراء مسح شامل للمجتمع، خاصة إذا كان ذلك المجتمع كبيرا جدا مثل سكان الصين، لذلك يكتفى بدراسة عينة من ذلك المجتمع. والسؤال الذي ينشأ الآن هو: كيف نختار عينة من مجتمع إحصائي؟

3 : 7 طرق اختيار العينة Sampling Techniques

إن جمع البيانات يتم بإجراء المسح الشامل أو بطريقة العينة وفي هذا البند سنتعرف على الطرق التي نختار بها العينة، ونبدأ ببعض التعاريف لكي نعطي العبارات معانيها الدقيقة:

تعريف (4)

مجتمع الهدف Target Population : هو كل المجتمع الذي نطلب المعلومات عنه فمثلا: جميع الطلبة على مقاعد الدراسة في الأردن من عمر 7 إلى 18 سنة.

أو: جميع الأسر في بلد معين.

أو: جميع المصابيح الكهربائية التي يصنعها مصنع في عام معين.

تعريف (5)

مجتمع الدراسة Study Population: هو مجموعة الأفراد التي يتاح لنا إجراء الدراسة عليها.

فمثلا: إذا كان مجتمع الهدف هو جميع الطلبة على مقاعد الدراسة في الأردن من عمر 7 إلى 18 سنة، فمن الممكن أن يكون مجتمع الدراسة جميع الطلبة في عمان من عمر 7 إلى 18 سنة.

تعريف (6)

سمة المجتمع Population Characteristic: هي الخاصية أو الوجه الذي نريد قياسه لدى المجتمع. فمثلا: إذا أراد طبيب الوقوف على حالة سلامة الأسنان لدى الأطفال من عمر 6 إلى 12 عاما في المملكة الأردنية الهاشمية، فإن سمة المجتمع هنا هي: "الحالة الصحية لأسنان الطفل". ومن الواضح أن

سمة المجتمع هذه يمكن قياسها بأحد تدريجات القياس التي مرّ ذكرها. إن قياس السمة يتغير من فرد إلى آخر، ولذلك فهو متغير. ونحن نحصل على معلومات عن سمة المجتمع من دراستنا لقياسات السمة لدى الأفراد جميعهم أو عينة منهم.

الهدف الرئيس هنا هو اختيار عينّة تمثل المجتمع وتؤدي إلى إحراز معلومات عن سمة المجتمع بشكل دقيق يتناسب مع التكلفة والجهد المستعملين: وهناك نوعان من العينات.

أولاً: العينات غير الاحتمالية Non - Probabilistic Sampling

نورد فيما يلي، أهم أنواع العينات غير الاحتمالية :

(1) المعاينة بالاختيار السهل Accessibility Sampling

وهي اختيار العينة باعتبار الأولوية لسهولة الوصول والحصول على المشاهدات، وفي هذه الحالة نختار العينة التي نستطيع أن نسجل القياسات والمشاهدات عن أفرادها بسهولة. فعلى سبيل المثال: إذا أردنا دراسة عادات التدخين بين طلبة الجامعة فإننا نطلب متطوعين للإجابة عن الأسئلة التي ستطرح. ومن سلبيات هذه الطريقة عدم كون العينة ممثلة للمجتمع بصورة واضحة جليّة.

ومثال آخر على هذه الطريقة : أراد مراقب السوق المركزي للخضار أن يعطي تقريرا عن جودة صناديق الخيار في السوق عن طريق دراسة عينة فكان هذا المراقب كلما دخلت شاحنة إلى السوق يأخذ أول صندوق من أعلى الشاحنة ويأخذ من سطحه عددا من الخيار ويتفحصها. من الواضح أن هذه العينة لا تمثل المجتمع تمثيلا صادقا.

(2) المعاينة الهادفة أو الحكمية: Purposive or Judgemental Sampling

والنظرة هنا مختلفة عن سابقتها، إذ أن الباحث في هذه الحالة يختار عينة بناء على حكم شخصي ورأي ذاتي، إذ يعتبر العينة المختارة ممثلة للمجتمع، مع علمه أن المجتمع يحتوى على أنواع مختلفة من الأفراد ذوي مقاييس مختلفة وتباين في سهولة الوصول إليهم.

ومن الممكن أن تغطي هذه الطريقة نتائج جيدة إذا كان رأي الباحث وحكمه صائبين.

وتحمل المعاينة الهادفة، في طياتها إلغاء مصادر التحريف المتوقعة، لكنه لا مناص من وجود التحريف، وذلك إما بسبب التحيز الشخصي، أو بسبب الجهل في جوانب بعض صفات المجتمع، وخاصة عند وجود ارتباط غير مكتشف بين طريقة المعاينة والمقياس الذي نريد دراسته.

ثانياً: العينات الاحتمالية Probability Sampling

نورد فيما يلي أهم أنواع العينات الاحتمالية:

(1) العينة العشوائية البسيطة: Simple Random Sample

إن هدف دراسة العينة هو التوصل إلى استقراءات واستنتاجات عن مجتمع الهدف مبنية على معلومات في العينة. وهناك كميتان من المعلومات تحتويهما العينة، وبالتالي فهما يؤثران في دقة عملية الاستقراء التي نتبناها. الأولى منهما هي حجم العينة المختارة من المجتمع، والثانية مقدار التغير في البيانات، ويمكن التحكم في هذا عادة بطريقة اختيار العينة.

وتسمى الطريقة التي نختار بها العينة "تصميم إجراء العينة" وسنبحث بعض تصاميم اختيار عينة ذات حجم ثابت n بشيء من التفصيل.

تعريف (7):

إذا اخترنا عينة ذات حجم n من مجتمع حجمه N بحيث يكون لكل عينة ذات حجم n نفس إمكانية الاختيار مثل أي عينة أخرى ذات الحجم نفسه، تسمى هذه الطريقة "المعاينة العشوائية البسيطة"، وتسمى العينة الحاصلة "عينة عشوائية بسيطة".

تستعمل المعاينة العشوائية البسيطة للحصول على تقديرات لمعدلات المجتمع والمجموع الكلي للمجتمع، والنسب الخاصة به.

ليس سحب عينة عشوائية بسيطة من مجتمع الهدف عملاً تافهاً كما يبدو لأول وهلة. كيف نستطيع سحب عينة من مجتمع بحيث يكون لكل عينة ممكنة ذات حجم n نفس إمكانية السحب مثل أي عينة أخرى؟

يمكن الحصول على عينات عشوائية بسيطة باستعمال جداول الأعداد العشوائية. وتجد أحد هذه الجداول في الملحق وتتصف العينة العشوائية البسيطة بأن يكون لأي مجموعة جزئية من مجتمع الدراسة، وبحجم معين نفس الفرصة في أن تختار كعينة من ذلك المجتمع مثلها مثل أي مجموعة جزئية أخرى بالحجم نفسه. وبشكل خاص يكون لأي عنصر في المجتمع فرصة مساوية لفرصة أي عنصر آخر في أن يكون مختاراً في العينة.

كيفية اختيار العينة العشوائية البسيطة:

(أ) اعط كل عنصر من عناصر مجتمع الدراسة رقماً متسلسلاً من 1 إلى N حيث N هو حجم مجتمع الدراسة. اجعل هذه الأرقام مكونة من نفس العدد من المنازل فمثلاً، إذا كان حجم المجتمع 900 وأردنا اختيار عينة عشوائية بسيطة منه، فإننا نعطي أفراد ذلك المجتمع الأرقام التالية :

001، 002، ..... ، 497، 498، 499، .....، 900

لاحظ أن الرقم المتسلسل لكل فرد يتكون من ثلاث منازل.

(ب) استعمل جداول الأرقام العشوائية واقرأ منه عمودياً بحيث يكون عدد منازل كل عدد مساوياً لعدد منازل الأرقام المتسلسلة في (أ)، فإذا كان العدد الذي قرأته من الجدول أحد الأرقام المتسلسلة فاعتبره عنصراً من عناصر العينة، وإلا فاتركه وانتقل إلى عدد آخر، ولا تكرر الأعداد التي أخذتها عناصر في العينة ومعنى ذلك إنك ترفض، أي عدد أخذته في قراءة سابقة.

واستمر في القراءة وكلما انتهيت من عمود انتقلت إلى العمود الذي يليه حتى تحصل على عينة بالحجم المطلوب.

مثال (1) :

عدد الطلبة المسجلين في مقرر الإحصاء 700 طالب والمطلوب اختيار عينة عشوائية بسيطة حجمها 10 من هؤلاء الطلبة.

الحل:
نعطي الطلبة أرقاماً متسلسلة، تبدأ من 1 وتنتهي بالعدد 700 وبذلك تكون أرقام الطلاب 700، 969، 697, ... ، 002، 001 تختار صفحة من جدول الأعداد العشوائية، ثم تبدأ بالعمود الأول من اليسار وتقرأ الأعداد واحداً تلو الآخر، بحيث تنظر إلى ثلاث منازل من جهة اليسار. فإذا كان العدد الذي تقرؤه واقعا ضمن أرقام الطلبة فانك تأخذ ذلك الرقم في العينة وإلا فاقرأ العدد الذي يليه, وهكذا.

نضع أمامك جزءا من جدول الأعداد العشوائية مؤلفاً من 20 عدداً لنشرح طريقة القراءة.

جزء من جدول الأعداد العشوائية

0695

0437

6242

7090

0683

7013

8808

9876

1873

2581

3785

8626

0113

4646

7873

3755

2673

0187

7976

والآن نبدأ بالقراءة من جدول الأعداد العشوائية فنقرأ ثلاث منازل من كل عدد ونأخذ العدد الواقع ضمن أرقام أفراد المجتمع ونرفض (لا نأخذ) أي عدد خارج عن حجم المجتمع ولا نكرر أي عدد أخذناه، فتكون العينة هي : 069، 043، 624، (709 لا نأخذه لأنه ليس في المجتمع)، 068، (701، 880، 987 لا نأخذها)، 187، 258، 378، (862 لا نأخذه) 011، 464، (787 لا نأخذه) آخر عدد نأخذه 375.

إن الطلبة الذين أرقامهم مذكورة أعلاه يؤلفون العينة المطلوبة وهي عينة عشوائية بسيطة حجمها 10.

(2) العينة الطبقية العشوائية: Stratified Random Sampling

هدف تصميم إجراء العينة في هذه الحالة هو الحصول على أعظم كميات من المعلومات بتكلفة معطاة، والواقع أن طريقة العينة العشوائية البسيطة، وهي تصميم أساس في المعاينة، تؤدي في كثير من الأحيان إلى تقديرات جيدة لسمات المجتمع بتكاليف منخفضة. سنتعرف الآن على طريقة ثانية للمعاينة نحتاجها عندما يكون المجتمع منقسما إلى طبقات طبيعية وتكون لدينا الرغبة في تمثيل جميع هذه الطبقات في العينة تؤدي إلى زيادة في المعلومات عند إعطاء تكلفة معيّنة.

تعريف (8)

عند تجزئة المجتمع إلى مجموعات غير متداخلة (تسمى طبقات) ومن ثم اختيار عينة عشوائية بسيطة من كل مجموعة فإننا نسمي هذه الطريقة: العينة الطبقية العشوائية.

كيفية اختيار عينة طبقية عشوائية:

(أ) أول خطوة في المعاينة الطبقية العشوائية هي تعيين الطبقات بوضوح، وبعد ذلك يأتي وضع كل وحدة معاينة من المجتمع في الطبقة الملائمة.

قد لا تكون هذه العملية بسيطة كما تبدو، فمثلا: إذا أردت أن تحدد الطبقات التي تنتمي إليها الأسر في وحدات حضرية وأخرى ريفية، فماذا تقرر عن قرية فيها 1000 نسمة؟ هل هذه القرية ريفية أم حضرية؟ يمكن أن نعتبرها ريفية إذا كانت بعيدة ومنفصلة عن مدينة كبيرة، أما إذا كانت بجوار مدينة كبيرة وقريبة منها فيمكن اعتبارها حضرية.

من ثم أصبح ضروريا أن تحدد ماذا تعني بكلمة: حضر، وكلمة ريف، لكي تستطيع وضع كل وحدة معاينة ضمن طبقة واحدة فقط.

(ب) بعد تقسيم وحدات المعاينة في طبقات وبالتالي تحديد حجم كل طبقة بعد ذلك نحدد حجم العينة من كل طبقة ومن ثم نختار عينة عشوائية بسيطة من كل طبقة باستعمال الطرق التي تم شرحها سابقا.

ويجب أن تتأكد أن العينات المختارة من الطبقات عشوائية ومستقلة عن بعضها البعض بمعنى أنه يجب استعمال طرق معاينة عشوائية في كل طبقة على حده، لكي نضمن أن المشاهدات التي نحصل عليها من طبقة معينة لا تعتمد على المشاهدات المختارة من طبقة أخرى.

تقسيم العينة على الطبقات بطريقة النسبة :

إن الغرض من تصميم العينة هو الحصول على عينة تمثل المجتمع. فإذا كان لديك k من الطبقات وأردت اختيار عينة حجمها n من المجتمع الكلي، فهناك عدة طرق لتقسيم الحجم n على الطبقات ، وسندرس منها طريقة النسبة.

افرض :

n_1: حجم العينة من الطبقة الأولى.

n_2: حجم العينة من الطبقة الثانية.

.

.

.

وهكذا إلى أن نصل إلى :

n_k: حجم العينة من الطبقة ذات الرقم k.

بحيث يكون $n_1 + n2 + ... + n_k = n$ أي أن حجم العينة الكلي يساوي مجموع حجوم العيّنات من الطبقات جميعها. وتنص طريقة النسبة على أن تحديد حجم العينة من أي طبقة بحيث يتناسب مع حجم الطبقة. فإذا كان حجم المجتمع N وكانت حجوم الطبقات كما يلي:

N_1 حجم الطبقة الأولى.

N_2 حجم الطبقة الثانية.

.

.

.

N_kحجم الطبقة التي رقمها k

بحيث يكون $N = N_1 + N_2 + ... + N_k$

وإذا كان حجم العينة المطلوب n فإننا نحّدد حجم العينة من كل طبقة بطريقة النسبة كما يلي:

$$n_1 = N_1 \times \frac{n}{N}$$

$$n_2 = N_2 \times \frac{n}{N}$$

$$n_k = N_k \times \frac{n}{N}$$

مثال (2) :
إذا كان طلاب السنة الأولى في إحدى الجامعات موزعين حسب الكليات كما يلي:

الآداب	العلوم	الاقتصاد	الهندسة
1200	1000	800	300

المطلوب اختيار عينة حجمها20% من المجتمع بطريقة المعاينة الطبقية العشوائية.

الحل:

حجم المجتمع: 300 + 800 + 1000 + 1200 = 3300

حجم العينة 660 = $3300 \times \frac{20}{100}$

من الواضح أن الطبقات هنا هي الكليات. ولما كان حجم العينة 20% من المجتمع الكلي، ونحن نريد أن تكون جميع الطبقات ممثلة بنفس النسبة فإن:

حجم العينة من كلية الآداب $1200 \times \frac{20}{100}$ 240 =

حجم العينة من كلية العلوم $1000 \times \frac{20}{100}$ 200 =

ومن كلية الاقتصاد $160800 \times \frac{20}{100}$ =

ومن كلية الهندسة $300 \times \frac{20}{100}$ 60 =

أولاً : نختار عينة عشوائية حجمها 240 من كلية الآداب التي فيها 1200 طالب وذلك باستعمال جدول الأعداد العشوائية كما عملنا في المثال (1).

ثانياً : نختار عينة عشوائية حجمها 200 من طلبة كلية العلوم.

ثالثاً : نختار عينة عشوائية حجمها 160 من طلبة الاقتصاد.

رابعاً : نختار عينة عشوائية حجمها 60 من طلبة كلية الهندسة.

مثال (3) :

إذا كانت طبقات أحد المجتمعات تحوى العناصر كما في الجدول التالي:

الطبقة الأولى	الطبقة الثانية	الطبقة الثالثة	الطبقة الرابعة	الطبقة الخامسة
500	400	280	200	220

وأراد باحث اختيار عينة حجمها 160 من هذا المجتمع. فما حجم العينة من كل طبقة؟

الحل: حجم المجتمع الكلي N يساوي

$500 + 400 + 280 + 200 + 220 = 1600$

حجم العينة الكلية $n = 160$

حجم العينة من الطبقة الأولى: $n_1 = 500 \times \frac{160}{1600} = 50$

حجم العينة من الطبقة الثانية: $n_2 = 400 \times \frac{160}{1600} = 40$

حجم العينة من الطبقة الثالثة: $n_3 = 280 \times \frac{160}{1600} = 28$

حجم العينة من الطبقة الرابعة: $n_4 = 200 \times \frac{160}{1600} = 20$

حجم العينة من الطبقة الخامسة: $n_5 = 220 \times \frac{160}{1600} = 22$

(3) العينة المنتظمة Systematic Sampling

تعريف (9):

إذا اخترنا عنصرا بطريقة عشوائية من أول k من العناصر في إطار المعاينة ومن ثم اخترنا كل عنصر رقمه k بعد العنصر المختار سابقا فإن هذه الطريقة تسمى طريقة العينة المنتظمة.

تعتبر المعاينة المنتظمة بديلا مفيدا للمعاينة العشوائية البسيطة للأسباب التالية:

(i) المعاينة المنتظمة أكثر سهولة في التنفيذ، ولذلك فهي أقل تعرضا لحصول أخطاء في المقابلات من المعاينة العشوائية البسيطة.

(ii) غالبا ما تعطي المعاينة المنتظمة معلومات لكل وحدة كلفة أكثر مما تفعل المعاينة العشوائية البسيطة.

ومن جهة عامة فإن المعاينة المنتظمة تتطلب اختيار عنصر بطريقة عشوائية من أول k من العناصر، وبعد ذلك نختار كل عنصر رقمه k بعد سابقه الذي تم اختياره. وهذه عملية أكثر سهولة وأقل تعرضا للأخطاء مقارنة بطريقة المعاينة العشوائية البسيطة.

مثال (4) :

إذا أردت أخذ عينة حجمها 20 من زبائن محل تجاري كبير فإنه يصعب عليك أن تحدد مجتمع الدراسة، وبالتالي يصعب إجراء معاينة عشوائية بسيطة لكنه من السهل بمكان أن تستعمل المعاينة المنتظمة، وذلك بأن تختار رقما بطريقة عشوائية من 1 إلى 10 مثلا. وليكن العدد الذي اخترته 6 فيعتبر هذا هو العنصر الأول في العينة، وتقابل الشخص السادس الذي يخرج من المحل التجاري مثلا، ثم تقابل الأشخاص الذين تكون أرقام خروجهم 16، 26، 36، وهكذا إلى أن تحصل على عينة حجمها 20.

من الواضح أن هذه طريقة سهلة حتى على شخص غير متدرب. وغالبا ما تكون العينة المختارة بالطريقة المنتظمة موزعة بشكل أكثر تجانسا على جميع أفراد المجتمع، وبالتالي يمكن أن تعطي معلومات عن المجتمع أكثر مما تعطيه عينة مماثلة تمّ اختيارها بالطريقة العشوائية البسيطة.

مثال (5) :

إذا أردت اختيار عينة حجمها n = 200 من مجموعة من بطاقات التسجيل في إحدى الجامعات التي سجّل فيها N = 3000 طالباً وذلك لندرس عدد البطاقات التي فيها أخطاء.

أن طريقة المعاينة المنتظمة تقضي بأن يكون طول القفزة $\frac{3000}{200} = 15$ ولذلك نسحب رقماً عشوائياً من 01 إلى 15 وليكن 8 ونختار البطاقة ذات رقم 8 ، ومن ثم نضيف 15 ونسحب البطاقة ذات الرقم 23 ثم البطاقة 38 ثم 53، وهكذا. وآخر بطاقة نسحبها هي رقم 2993 كما في الجدول التالي:

رقم البطاقة	رقم البطاقة المختارة
1	
2	
3	
.	8
.	

	4 15
23	16 17 18 . . 30
38	31 32 33 . . 45

وهكذا إلى أن تصل إلى آخر خمس عشرة بطاقة (أرقامها 2986،2987، ... ،3000) فنختار البطاقة التي رقمها 2993.

ونلاحظ هنا أنه إذا لم يكن طول القفزة عددا صحيحا فإننا نقرب الجواب إلى أقرب عدد صحيح. ففي مثال (5) لو كان عدد الطلبة 3065 على سبيل المثال فإن الحل سيبقى كما هو ويكون طول القفزة $\mathbf{15 = \frac{3065}{200}}$ تقريبا.

وبالإشارة إلى المجتمع والعينة نجد أن الإحصاءات تخدم هدفين. إنها أولاً تصف العينة ذاتها، كأن نجد معدل العينة، أو مدى البيانات في العينة، أو نسبة النجاح في العينة، وثانيا تمكننا من التوصل إلى استقراءات واستنتاجات عن المجتمع الذي أخذنا منه العينة، وقد لخص ملتون وآخرون خطوات الدراسة الإحصائية كما يلي :

خطوات الدراسة الإحصائية Steps in a Statistical Study[1]

1- عرف وعين مجتمع الدراسة.

2- حدد الأسئلة التي تريد الإجابة عليها بما يتعلق بالمجتمع.

3- حدد المتغيرات ذات الاهتمام والتي تساعدك في الإجابة عن الأسئلة التي طرحتها.

4- حدد معلمات المجتمع التي تهمك.

5- اسحب عينة من المجتمع.

6- أوجد قيم الإحصاءات التي تعطيك قيما تقريبية عن معلمات المجتمع.

7- طبق طرق الإحصاء الاستقرائي للإجابة عن الأسئلة المطروحة.

7 : 4 إحصاءات العينة Sample Statistics

يخضع المجتمع الذي تؤخذ منه العينة لتوزيع معين هو توزيع المجتمع. وهو التوزيع الاحتمالي لمتغير عشوائي يمثل أفراد ذلك المجتمع. وكما درست سابقا فإنه يوجد في التوزيع الاحتمالي عادة ثوابت تعين هذا التوزيع تماما، وتسمى معلمات.

فمثلا إذا كان المجتمع يخضع لتوزيع بيرنوللي فإن المعلمة هي احتمال النجاح للمحاولة p. فإذا ما علمت p فإنك تستطيع الإجابة عن الأسئلة الاحتمالية حول هذا المجتمع. أي أن توزيعه يكون قد علم تماما. أما إذا كان المجتمع يخضع لتوزيع طبيعي فإن المعلمات هي الوسط والتباين، فإذا ما علمت قيمة الوسط μ وقيمة التباين σ^2 فإن المجتمع يتحدد تماما، أي أن اقتران الكثافة الاحتمالية للمتغير X الذي يمثل أي فرد في ذلك المجتمع يكون قد تحدد تماما. ويصبح بإمكاننا الإجابة عن الأسئلة المتعلقة بالاحتمالات حول X. عندما تأخذ عينة من مجتمع ما ، فإن هذا يعني أنك ستشاهد قيما للمتغير X، وإذا كان حجم العينة n فإنك تعبر عن هذه القيم ب : $x_n, \ldots x_2, x_1$. ويمكن استعمال أي حرف غير X لتعبر عن المتغير قيد البحث.

وكما درسنا سابقا فيمكننا حساب بعض المقاييس عن هذه العينة مثل الوسط الحسابي أو الوسيط أو التباين وغيرها. هذه المقاييس تسمى إحصاءات.

وتستعمل كلمة "إحصاء" أو "إحصاء عينة" لتدل على أي اقتران تتعين قيمته من العينة.

---

[1] Milton, J.S., Corbet, J.J., Mc Teer, P.M. "Introduction to Statistics". Heath and Company 1986 Toronto.

فمثلا الوسط لحسابي للعينة : $\overline{X} = \sum_{i=1}^{n} X_i / n$ هو إحصاء عينة. ويلاحظ أن قيمته ربما تتغير من عيّنة لأخرى. أي إذا أخذت عينة $X_1, X_2, \dots, X_n$ فإن هذه العينة تحدد قيمة للوسط الحسابي، وإذا أخذت عينة أخرى بنفس الحجم فإن الوسط الحسابي لهذه العينة ربما يختلف عن الوسط الحسابي للعينة الأولى وإذا أخذت عينة ثالثة فربما تحصل على قيمة ثالثة للوسط الحسابي وهكذا. وهذا يعني أن $\overline{X}$ متغير عشوائي، أي أن قيمة هذا الإحصاء تتغير بتغير العينة.

تعريف (10) :
إحصاء العينة هو متغير تتعين قيمه من جميع العينات ذات حجم معين مأخوذة من مجتمع ما.
وباختصار هو اقتران تتعين قيمته من العينة.

والسؤال الذي ينشأ الآن : إذا أخذنا كل العينات العشوائية ذات الحجم n من مجتمع إحصائي، فما نوع التوزيع الذي يخضع له الإحصاء $\overline{X}$.

تعريف (11) :
يسمى التوزيع الاحتمالي لإحصاء العينة توزيع المعاينة لذلك الإحصاء.

7 : 5 توزيع المعاينة للوسط الحسابي $\overline{x}$:

Sampling Distribution of the Mean $\overline{X}$

قبل تحديد توزيع المعاينة للوسط الحسابي للعينة $\overline{X}$ ، نحتاج إلى دراسة بعض المعلمات التي يمكن أن نحتاجها لوصف ذلك التوزيع، ومنها المعدل والتباين وعلاقتها بمعدل المجمتع الأم وتباينه، وهذا سيتم شرحه في المثال التالي :

مثال (6) :
مجتمع إحصائي حجمه N = 6 وهي 11 ,8 ,7 ,6 ,3 ,1

نريد سحب جميع العينات ذات الحجم n = 2 مع الارجاع.

(يمكن النظر إلى هذا المجتمع على أنه مؤلف من 6 بطاقات كل بطاقة مكتوب عليها قيمة من القيم أعلاه ونريد سحب العينات ذات حجم n = 2 مع الارجاع).

أ- احسب معدل المجتمع وتباينه.

ب- سجل جميع العينات الممكنة ذات الحجم n = 2 مع الارجاع واحسب الوسط الحسابي لكل منها.

جـ- ضع نتائج (ب) في توزيع تكراري.

د- احسب الوسط الحسابي والتباين للتوزيع في (جـ).

هـ- قارن نتائجك في (د) مع النتائج في (أ).

لاحظ أن $\overline{X}$ متغير عشوائي تعتمد قيمته على العينة التي تختارها ولذلك فله وسط (توقع رياضي) وتباين وتوزيع احتمالي ,

أ- الاقتران الاحتمالي هو : $f(x)=\frac{1}{6}\ ,\ \ x=1,3,6,7,8,11$

معدل المجتمع $\mu=\sum x.f(x)$

$$=1\times\frac{1}{6}+3\times\frac{1}{6}+6\times\frac{1}{6}+7\times\frac{1}{6}+8\times\frac{1}{6}+11\times\frac{1}{6}$$
$$=6$$

تباين المجتمع هو :

$$\sigma^2=\sum(x-\mu)^2.f(x)$$

ب-

$$=\frac{1}{6}\left[(1-6)^2+(3-6)^2+(6-6)^2+(7-6)^2+(8-6)^2+(11-6)^2\right]$$

$$=\frac{1}{6}[25+9+0+1+4+25]=\frac{64}{6}=\frac{32}{3}$$

نرتّب الحل كما في الجدول :

العينة	وسطها الحسابي	العينة	وسطها الحسابي	العينة	وسطها الحسابي
(1 ,1)	1	(6 ,1)	3.5	(8 ,1)	4.5
(1 ،3)	2	(6 ،3)	4.5	(8 ,3)	5.5
(1 ،6)	3.5	(6 ،6)	6	(8 ,6)	7
(1 ،7)	4	(6 ,7)	6.5	(8 ،7)	7.5
(1 ،8)	4.5	(6 ,8)	7	(8 ,8)	8

(1 ،11)	6	(6 ,11)	8.5	(8 ,11)	9.5
(3 ،1)	2	(7 ,1)	4	(11 ,1)	6
(3 ،3)	3	(7 ,3)	5	(11 ,3)	7
(3 ،6)	4.5	(7 ,6)	6.5	(11 ,6)	8.5
(3 ،7)	5	(7 ,7)	7	(11 ,7)	9
(3 ،8)	5.5	(7 ,8)	7.5	(11 ,8)	9.5
(3 ،11)	7	(7 ,11)	9	(11,11)	11

جـ- التوزيع التكراري للأوساط الحسابية هو العمودان (1) ، (2) في الجدول التالي :

(1) $\bar{X}_i$	(2) التكرار f_i	(3) $\bar{X}_i f_i$	(4) $\bar{X}_i^2$	(5) $\bar{X}_i^2 f_i$
1	1	1.0	1	1
2	1	4.0	4	8
3	1	3.0	9	9
3.5	2	7.0	12.25	24.5
4	2	8.0	16.0	32.0
4.5	4	18.0	20.25	81.0
5	2	10.0	25	50
5.5	2	11.0	30.25	60.5
6	3	18	36.00	108.0
6.5	2	13	42.25	84.5
7	5	35	49.00	245.0

7.5	2	15	56.25	112.5
8	1	8	64.00	64.0
8.5	2	17	72.25	144.5
9	2	18	81.00	162.0
9.5	2	19	90.25	180.5
11	1	11	121.00	121.0
المجموع	36	216		1488

د- معدل الأوساط الحسابية هو $\mu_{\bar{x}} = \frac{\sum \bar{X}_i f_i}{m} = \frac{216}{36} = 6$

تباين الأوساط الحسابية هو :

$$\sigma_{\bar{x}}^2 = \frac{\sum \bar{X}_i^2 f_i}{m} - \mu_{\bar{x}}^2 = \frac{1488}{36} - 6^2 = 41.3 - 36 = 5.33$$

لاحظ أننا في حساب التباين $\sigma_{\bar{x}}^2$ قسمنا على مجموع التكرارات m لأن التوزيع التكراري يمثل مجتمع جميع العينات الممكنة ذات الحجم 2 مع الإرجاع أي أن هذا التوزيع التكراري يمثل مجتمعا ولا يمثل عينة ولذلك حسبنا $\sigma_{\bar{x}}^2$ ولم نحسب S^2.

ولدى مقارنة معدل $\bar{X}$ (التوقع الرياضي للوسط $\bar{X}$) مع التوقع الرياضي للمجتمع نجد أنهما متساويان أي أن $\mu_{\bar{X}} = 6 = \mu$ وعند مقارنة تباين $\bar{X}$ مع تباين المجتمع نجد أن

$$\sigma_{\bar{x}}^2 = 5.33 = \frac{32}{3 \times 2} = \frac{\sigma^2}{n}$$

إن هذا المثال هو تطبيق مباشر للنظرية التالية :

نظرية (1)

إذا كان X يخضع لتوزيع وسطه (معدله) μ وتباينه σ^2 ، وكان $\bar{X}$ يمثل الوسط الحسابي للعينة ذات الحجمn والمسحوبة من هذا المجتمع فإن :

(1) توقع $\overline{X}$ هو $\mu_{\bar{x}} = \mu$

(2) تباين $\overline{X}$ هو $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$

شريطة أن السحب مع الإرجاع أو المجتمع لا نهائي.

مثال (7) :

أخذت عينة مع الارجاع حجمها 3 مع المجتمع 10, 2, 4, 3, 6. فإذا كان $\overline{X}$ الوسط الحسابي للعينة أوجد توقع $\overline{X}$ وتباينه.

الحل : يمكننا إيجاد الحل كما في المثال (1) أما الآن فنطبق نظرية(1) لإيجاد الحل مباشرة.

نجد الوسط الحسابي للمجتمع : $\mu = \frac{6+3+4+2+10}{5} = 5$

ونجد تباين المجتمع :

$$\sigma^2 = \frac{(6-5)^2 + (3-5)^2 + (4-5)^2 + (2-5)^2 + (10-5)^2}{5}$$

$$= \frac{1+4+1+9+25}{5} = \frac{40}{5} = 8$$

افرض الوسط الحسابي للعينة $\overline{X}$ ، وبتطبيق النظرية (1) نجد :

$$\mu_{\bar{x}} = \mu = 5$$

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} = \frac{8}{3}$$

مثال (8) :

سحبت عينة عشوائية من مجتمع لا نهائي معدله 60 وتباينه 49. فإذا كان حجم العينة 20 فأوجد معدل الوسط الحسابي (معدل $\overline{X}$) وتباين $\overline{X}$ ، أوجد الانحراف المعياري للمعدل $\overline{X}$.

الحل : السحب من مجتمع لا نهائي ولذلك من نظرية (1) ، معدل $\overline{X}$ هو

$$\mu_{\bar{x}} = \mu = 60$$

وتباين $\overline{X}$ هو

$$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n} = \frac{49}{20} = 2.45$$

الانحراف المعياري للمعدل $\overline{X}$ (يساوي الجذر التربيعي للتباين) أي $\sigma_{\bar{x}} = \sqrt{2.45} = 1.56$.

مثال (9) :
سحبت عينة عشوائية حجمها n من مجتمع توزيعه طبيعي ذو معدل 25 وتباين 40 فإذا كان تباين الوسط الحسابي 8 فما هو حجم العينة ؟

الحل : من نظرية (1) : $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$

إذن $8 = \frac{40}{n}$

إذن 5 = n حجم العينة.

لاحظ أننا وجدنا معدل الوسط الحسابي $\overline{X}$ وتباينه ولكننا لم نتعرض لدراسة توزيع $\overline{X}$ وهذا ما سنفعله في البنود القادمة.

7 : 6 توزيـع المعاينـة للوسـط الحسـابي $\bar{x}$ عنـد المعاينـة مـن مجتمع طبيعي

Sampling Distribution of the Mean $\overline{X}$ when sampling from Normal Distribution.

إذا أخذت العينة العشوائية $X_1, X_2, \dots, X_n$ من مجتمع ما ، وكان $\overline{X}$ الوسط الحسابي لهذه العينة، فما هو توزيع $\overline{X}$ ؟

قبل الإجابة عن هذا السؤال عليك أن تلاحظ أن في مرحلة التخطيط لدراسة عينة ما ، وقبل جمع البيانات، أو تسجيل قيم المشاهدات، يكون بإمكانك التعبير عن الوسط الحسابي $\overline{X}$ بدلالة الاحتمالات فقط، لأن الوسط حينذاك يكون عبارة عن مجموع متغيرات عشوائية مقسوما على عدد ثابت n. ولذلك

فقبل إجراء التجربة وتسجيل قيم المشاهدات يكون الإحصاء $\overline{X}$ متغيرا عشوائيا، أما بعد إجراء التجربة فإن قيمته تتعين من العينة المشاهدة وتستطيع أن تعبر عن هذه القيمة بالحرف الصغير $\overline{x}$.

وبما أن $\overline{X}$ متغير عشوائي، فإنه يكون له توزيع احتمالي يعرف بتوزيع المعاينة للوسط، وهو التوزيع الاحتمالي لجميع الأوساط $\overline{X}$ لجميع العينات العشوائية البسيطة ذات حجم معين والتي يمكن أخذها من المجتمع تحت الدراسة.

والسؤال الآن هو إذا أخذت جميع العينات العشوائية البسيطة ذات الحجم n من مجتمع إحصائي فما هو التوزيع الذي يخضع له الإحصاء $\overline{X}$ ؟

النظرية التالية تجيب على هذا السؤال عندما يكون المجتمع الإحصائي المجتمع الطبيعي.

نظرية (2) :إذا كانت $X_1, X_2, ..., X_n$ عينة عشوائية من مجتمع طبيعي وسطه (معدله) μ ، وتباينه σ^2 فإن توزيع $\overline{X}$ يكون التوزيع الطبيعي ذا الوسط μ والتباين $\frac{\sigma^2}{n}$ أي أن المتغير العشوائي :

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

يخضع لتوزيع طبيعي معياري.

مثال (10) :يخضع طول حياة مصابيح كهربائية إلى توزيع طبيعي وسطه 700 ساعة وانحرافه المعياري 60 ساعة (أ) أوجد احتمال أن يكون معدل الحياة لعينة حجمها 25 مصباحا أقل من 670 ساعة؟ (ب) أوجد احتمال أن يكون معدل الحياة لهذه العينة أكثر من 718 ساعة.

الحل : حسب النظرية (2) يكون توزيع $\overline{X}$ ، الوسط الحسابي للعينة، توزيعا طبيعيا ذا وسط 700 وانحراف معياري $\frac{60}{\sqrt{25}} = 12$ ساعة.

أي أن $Z = \frac{\overline{X} - 700}{12}$ يخضع للتوزيع الطبيعي المعياري .

(أ) : $$P(\overline{X} < 670) = P\left(\frac{\overline{X} - 700}{12} < \frac{670 - 700}{12}\right)$$

$$= P(Z < -2.5) = 0.0062$$

(ب) :
$$P(\overline{X} > 718) = P\left(\frac{\overline{X} - 700}{12} > \frac{718 - 700}{12}\right)$$
$$= P(Z > 1.5) = 1 - 0.9332 = 0.0668$$

مثال (11) :

تخضع علامات الطلاب في أحد المقررات لتوزيع طبيعي وسطه 65 وانحرافه المعياري 18 أخذت عينة عشوائية حجمها 36. احسب احتمال أن يزيد وسط علامات العينة على 74؟

الحل : افرض أن وسط العينة المسحوبة هو $\overline{X}$ والمطلوب حساب ($\overline{X}$ >74)P وبتطبيق نظرية(2) نجد :

$$P(\overline{X} > 74) = P\left(\frac{\overline{X} - 65}{18/\sqrt{36}} > \frac{74 - 65}{18/6}\right)$$
$$= P(Z > 3)$$
$$= 1 - 0.9987 = 0.0013$$

7 : 7 توزيع المعاينة للوسط الحسابي $\overline{x}$ عند المعاينة من مجتمع غير طبيعي (نظرية النهاية المركزية) :

Sampling Distribution of the mean $\overline{X}$ when sampling from a non-normal Distribution (Central Limit Theorem)

تتلخص المسألة التي نحن بصدد بحثها الآن في ما يلي :

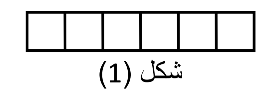
إذا كان لدينا مجتمع وسطه μ وتباينه σ^2، وأخذت منه جميع العينات العشوائية ذات الحجم n فما هو توزيع الوسط الحسابي $\overline{X}$ حتى لو لم يكن توزيع المجتمع التوزيع الطبيعي ؟

للإجابة عن هذا السؤال :

نعود إلى مثال (1) ونلاحظ أن قيم المجتمع متساوية إمكانية الحدوث وبالتالي يكون اقترانها الاحتمالي كما يلي :

القيمة	الاحتمال P(x)
1	1/6
3	1/6
6	1/6
7	1/6
8	1/6
11	1/6

وهو توزيع متجانس ومدرجه التكراري كما في الشكل (1) :



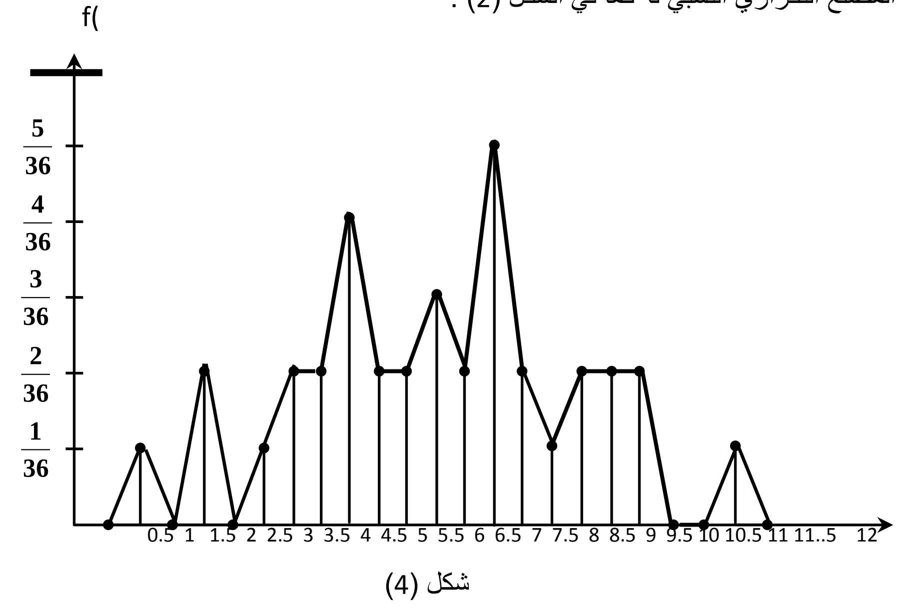
شكل (1)

أما التوزيع التكراري النسبي للوسط الحسابي $\overline{\mathbf{X}}$ لجميع العينات ذات الحجم n = 2 التي سحبناها مع الارجاع فنحصل عليه من المثال (6) فرع (حـ) وهو كما يلي :

القيمة	الاحتمال P(x)
1	1/36
2	2/36
3	1/36
3.5	2/36
4	2/36
4.5	4/36
5	2/36
5.5	2/36

6	3/36
6.5	2/36
7	5/36
7.5	2/36
8	1/36
8.5	2/36
9	2/36
9.5	2/36
11	1/36

المضلع التكراري النسبي له كما في الشكل (2) :



شكل (4)

وبالنظر إلى هذا المضلع نرى أنه بدأ يقرب من التوزيع الطبيعي.

لو أخذنا جميع العينات الممكنة ذات الحجم n = 3 مع الارجاع وحسبنا $\overline{X}$ لكل منها ثم وجدنا التوزيع التكراري النسبي للقيم $\overline{X}$ التي حصلنا عليها وبعد ذلك رسمنا المضلع التكراري النسبي لذلك التوزيع

لحصلنا على توزيع قريب من التوزيع الطبيعي. ولو كررنا هذه العملية لقيم مختلفة للحجم n ، أي n = 10 ، n = 15، وهكذا لوجدنا أن التوزيع التكراري النسبي للقيم $\overline{X}$ لجميع العينات الممكنة ذات الحجم n يقرب من التوزيع الطبيعي أكثر وأكثر.

قارن بين الشكل (1) والشكل (2). يمثل الشكل (1) منحنى توزيع المجتمع ، ويتضح أنه على شكل مستطيل. ويمثل الشكل (2) المضلع التكراري النسبي للأوساط الحسابية لـ 36 عينة، حجم كل منها 2 ويتضح أن شكل هذا المضلع يقرب من شكل منحنى التوزيع الطبيعي. وعند حساب الوسط الحسابي للتوزيع كما في (جـ) تجد أنه يساوي الوسط الحسابي للمجتمع. وعند حساب تباين هذا التوزيع تجد "أنه يقرب" من تباين المجتمع الأصلي مقسوما على حجم العينة.

نظرية (3) :

نظرية النهاية المركزية (تقارب التوزيعات) Central Limit Theorem

إذا أخذت عينة عشوائية حجمها n من مجتمع إحصائي وسطه μ وتباينه σ^2.

فإن توزيع الوسط الحسابي للعينة $\overline{X}$ يقترب من التوزيع الطبيعي ذي الوسط الحسابي μ والتباين σ^2/n كلما كبرت n.

وبعبارة أخرى، فإن $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ يقترب من التوزيع الطبيعي المعياري كلما كبرت n وعند التطبيق $n \geq 30$ تعتبر كبيرة بشكل كاف يمكنك من استعمال هذه النظرية.

مثال (12) :

تخضع أوزان علب سائل غسل الصحون من نوع معين لتوزيع معدله 1000غم وانحرافه المعياري 20 غم.

إذا كانت هذه العلب تعبأ في صناديق وكان كل صندوق يحوي 36 علبة، أوجد

(أ) احتمال أن وزن صندوق أخذ عشوائيا ينقص عن 35.28 كغم.

(ب) احتمال أن وزان صندوق أخذ عشوائيا يزيد على 36.18 كغم.

الحل : إذا افترضنا أوزان العلب في الصندوق هي $X_1, X_2, \ldots, X_{36}$ يكون وزن الصندوق $\sum_{1}^{36} X_i$ إذاً فالمطلوب :

(أ) $P\left(\sum_{i=1}^{36} X_i < 35280\right)$ حيث حوّلنا كغم إلى غرامات

أي $P\left(\overline{X} < \frac{35280}{36}\right)$ أي $P(\overline{X} < 980)$

وبما أن حجم العينة 36 يمكننا استعمال نظرية النهاية المركزية ويكون :

$$P(\overline{X} < 980) = P\left(\frac{\overline{X} - 1000}{20/\sqrt{36}} < \frac{980 - 1000}{20/\sqrt{36}}\right)$$
$$= P(X < -6) = 0$$

وبالمثل فالمطلوب في (ب) هو :

$$P\left(\sum X_i > 36180\right) = P\left(\overline{X} > \frac{36180}{36}\right)$$
$$= P(\overline{X} > 1005)$$
$$= P\left(\frac{\overline{X} - 1000}{20/\sqrt{36}} > \frac{1005 - 1000}{20/\sqrt{36}}\right)$$
$$= P(Z > 1.5) = 1 - 0.9332 = 0.0668$$

بتطبيق نظرية النهاية المركزية.

مثال (13) :

تخضع أوزان عبوات أحد مبيدات الحشرات المنزلية لتوزيع وسطه 135غم وانحرافه المعياري 14غم. إذا قررت وزارة التموين رفض كل صندوق من هذه العبوات إذا نقص وزنه عن 6.24كغم، فما نسبة الصناديق المرفوضة، علما بأن عدد العبوات في كل صندوق 48 عبوة ؟

الحل : المطلوب حساب احتمال أن ينقص وزن الصندوق عن 6.24 كغم. فإذا افترضنا أوزان العلب في صندوق ما هي : $X_1, X_2, ..., X_{48}$ فالمطلوب إيجاد :

$$P\left(\sum_{i=1}^{48} X_i < 6240\right) = P\left(\frac{\sum X_i}{48} < \frac{6240}{48}\right)$$
$$= P(\bar{X} < 130)$$

ولما كانت شروط النظرية متحققة، حيث μ, σ^2 معلومان ، وحجم العينة أكبر من 30، فإن بإمكاننا تطبيق تلك النظرية باعتبار $\frac{\bar{X} - 135}{14\sqrt{48}}$ يخضع لتوزيع طبيعي معياري.

إذا نحسب $P(\bar{X} < 130) =$

$$P\left(\frac{\bar{X} - 130}{14/\sqrt{48}} < \frac{130 - 135}{14/\sqrt{48}}\right) = P(Z < -2.47)$$
$$= 0.0068 \approx 0.007$$

أي أن الوزارة ترفض 7 بالآلف من الصناديق تقريبا. من الواضح في الأمثلة التي درسناها أن تباين المجتمع σ^2 كان معلوما ولذلك أمكننا تطبيق النظريات المتعلقة بالوسط $\bar{X}$ في حالة المعاينة من مجتمع طبيعي أو الحالة التي فيها حجم العينة كبيرا.

والسؤال الذي يطرح نفسه هو : ماذا يحصل لو لم يكن التباين σ^2 معلوما ؟

أي أنه عند تطبيق النظريات السابقة وفي حالة σ^2 غير معلومة فماذا نعوض عنها؟

يكون الحال سهلا عندما يكون حجم العينة كبيرا أي عندما $n \geq 30$ حيث يمكننا استعمال S^2، تباين العينة بدلا من σ^2.

أما في حالة الحجوم الصغيرة أي $n < 30$ فإننا سندرس توزع $\bar{X}$ في البند التالي.

7 : 8 المعاينة من مجتمع طبيعي وسطه μ وتباينه σ^2 غير معلوم.

Sampling from Normal Distribution with unknown variance

في كثير من التطبيقات المتعلقة باستعمال الوسط الحسابي $\bar{X}$ لا يكون الانحراف المعياري للمجتمع σ معروفا، ولذلك إذا كان حجم العينة n يساوي 30 أو أكثر فإننا نستعمل S (الانحراف المعياري للعينة)

بدلا من σ ، ويكون $\frac{\overline{X}-\mu}{S/\sqrt{n}}$ خاضعا لتوزيع يمكننا تقريبه بالتوزيع الطبيعي المعياري ما دامت n ≤30، ولكن في الحالات التي يكون فيها n أقل من 30 فإن الانحرافات المعيارية للعينة تكون ذات تغير كبير لدرجة أن S لا تكون تقديرا موثوقا للانحراف المعياري σ وبالتالي $\frac{\overline{X}-\mu}{S/\sqrt{n}}$ يخضع لتوزيع لا يمكننا تقريبه بالتوزيع الطبيعي المعياري.

نظرية (4) :

إذا أخذت عينات عشوائية من توزيع طبيعي وسطه μ وتباينه σ^2 غير معلوم.

وإذا كان $\overline{X}$ الوسط الحسابي لعينة حجمها n وكان S الانحراف المعياري لهذه العينة، فإن المتغير :

$$T=\frac{\overline{X}-\mu}{S/\sqrt{n}}$$

يخضع لتوزيع t على درجات حرية m = n - 1

7 : 9 توزيع الفرق بين وسطين :

Distribution of the difference between two sample means

(1) التوزيعات الطبيعية المستقلة :

إذا أخذت العينات العشوائية من توزيعات طبيعية مستقلة عن بعضها فإن توزيع الفرق بين الوسطين يعطى بالنظرية التالية :

نظرية (5) :

إذا أخذت عينة عشوائية حجمها n_1 من توزيع طبيعي معدله μ_1 وتباينه σ_1^2 وأخذت عينة عشوائية حجمها n_2 من مجتمع طبيعي معدله μ_2 وتباينه σ_2^2 ومستقل عن المجتمع الأول ، ورمزنا للوسط الحسابي للعينة الأولى بالرمز $\overline{X}$ وللوسط الحسابي للعينة الثانية $\overline{Y}$ فإن توزيع $\left(\overline{X}-\overline{Y}\right)$ يكون التوزيع الطبيعي ذا المعدل $\left(\mu_1-\mu_2\right)$ والتباين $\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}$ ، أي أن :

$$Z=\frac{\left(\overline{X}-\overline{Y}\right)-\left(\mu_1-\mu_2\right)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}$$

يخضع لتوزيع طبيعي معياري N (0, 1).

مثال (14) :
إذا كانت رواتب المعلمين في وزارة التربية والتعليم تخضع لتوزيع طبيعي معدله 230 دينارا وانحرافه المعياري 36 دينارا ورواتب المعلمين في المدارس الخاصة تخضع لتوزيع طبيعي معدله 180 دينارا وانحرافه المعياري 40 دينارا.

أخذت عينة عشوائية من المعلمين في الوزارة حجمها 16 معلما وعبرنا عن وسطها الحسابي بالرمز $\overline{X}$ وأخذت عينة عشوائية من معلمي المدارس الخاصة حجمها 10 وعبرنا عن وسطها الحسابي بالرمز $\overline{Y}$. أوجد احتمال أن يزيد $\overline{X}$ عن $\overline{Y}$ بمقدار 60.

الحل : نريد إيجاد $P(\overline{X}-\overline{Y}\geq 60)$

بتطبيق النظرية (5) :

$$P(\overline{X}-\overline{Y}\geq 60)=P\left(\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\geq\frac{60-(230-180)}{\sqrt{\frac{(36)^2}{15}+\frac{(40)^2}{10}}}\right)$$

$$=P\left(Z>\frac{10}{\sqrt{81+160}}\right)$$

$$=P(Z>0.645)$$

$$=1-P(Z\leq 0.645)=1-0.7389=0.2611$$

7 : 10 توزيع المعاينة للنسبة والفرق بين نسبتين :

Sampling Distribution of Proportion and the difference between two proportions.

إذا كانت قيمة كل عنصر في مجتمع 0 أو 1 أي "لا" أو "نعم" أي "فشل" أو "نجاح" فإننا نسمي كل مشاهدة من هذا المجتمع "تجربة بيرنوللي".

وعندما نأخذ عينة عشوائية من هذا المجتمع فإن المعاينة في هذه الحالة تكون من مجتمع بيرنوللي.

وينصب اهتمامنا في هذه الحالة على نسبة النجاح في العينة.

تختلف نسبة النجاح في العينة من عينة إلى أخرى. فلو كان احتمال النجاح (الحصول على 1) في أي تجربة من مجتمع بيرنوللي يساوي p ، وأخذت عينة عشوائية حجمها n من هذا المجتمع وعبرت عن عدد (النجاح) بالرمز X فإن نسبة النجاح في العينة تكون : $\overline{P} = \frac{X}{n}$

إذا أخذت عينة عشوائية أخرى من هذا المجتمع ترى أن نسبة النجاح في العينة تتغير ، ولذلك فمن الواضح أن نسبة النجاح في العينة $\overline{P}$ متغير عشوائي، وللتعرف على موضوع توزيع نسبة النجاح في العينة، ندرس المثال التالي :

1- يخضع إنتاج مصنع مصابيح كهربائية للفحص الدوري ، فإذا وقعت أطوال حياة المصابيح ضمن حدود معينة اعتبرت المصابيح صالحة للاستعمال. وإذا خرجت عن تلك الحدود اعتبرت المصابيح غير صالحة للاستعمال. فإذا فحصت n من المصابيح ووجدت X منها صالحة فإن نسبة الإنتاج الصالح تكون $\overline{P} = \frac{X}{n}$.

إذا كررت هذه التجربة عدة مرات تغيرت قيمة $\overline{P}$ ، وهذا يعني أن $\overline{P}$ متغير عشوائي، والسؤال هنا : ما هو توزيع الإحصاء $\overline{P}$ وما هي خصائصه ؟

للإجابة عن هذه الأسئلة قارن $\overline{P}$ بالوسط الحسابي للعينة ذات الحجم n المأخوذة من مجتمع قيمة كل فرد فيه إما 1 أو 0. ولتوضيح ذلك عرف المتغير العشوائي X_i على النحو التالي :

إذا كان المصباح الذي رقمه أ صالحا تضع $X_i = 1$.

إذا كان المصباح الذي رقمه i غير صالح تضع $X_i = 0$.

حيث $i = 1, 2, \dots , n$

لاحظ أن $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ ويساوي (عدد المصابيح الصالحة)/ n وهذا هو $\overline{P}$.

ومن هنا فإن نسبة النجاح $\overline{P}$ هي وسط حسابي لعينة ذات حجم n مأخوذة من مجتمع بيرنوللي أي ذات الحدين الذي معلمتاه (1, p) والذي يأخذ القيمة 1باحتمال p. والقيمة 0 باحتمال (p- 1) .

لاحظ أن معدل مجتمع ذات الحدين (p ، 1) هو p = 1. p وتباينه يساوي (p(1-p .1، وأن التوقع للوسط الحسابي يساوي معدل المجتمع والتباين للوسط الحسابي يساوي تباين المجتمع مقسوما على حجم العينة، لذلك فإن توقع $\overline{P}$ يكون p وتباين $\overline{P}$ يكون $\frac{p(1-p)}{n}$.

وبتطبيق نظرية النهاية المركزية يقترب توزيع $\overline{P}$ من التوزيع الطبيعي عندما يزداد الحجم n. ونوجز هذا في النظرية :

نظرية (6) :

إذا أخذت عينة عشوائية حجمها n من مجتمع إحصائي يخضع لتوزيع بيرنوللي أي ذات الحدين (,1 p). وكانت نسبة النجاح المتوفرة من العينة هي $\overline{P}$ فإن توزيع $\overline{P}$ يقرب من التوزيع الطبيعي الذي وسطه p وتباينه $\frac{p(1-p)}{n}$ كلما كبرت n (أي $n \geq 30$).

لاحظ أنه عندما نقول احتمال النجاح في مجتمع ذات الحدين هو p فإن كلمة "النجاح" هنا تعني "حدوث الظاهرة التي اصطلحنا أن حدوثها هو نجاح "فربما أن "النجاح" تعني "حدوث الوفاة في عملية جراحية" ، وربما تعني "استعمال نظارة طبية"، وربما "الحصول على درجة البكالوريوس بتقدير جيد جدا"وهكذا، فإن p تفسر على أنها نسبة الأفراد الذين لهم صفة معينة في مجتمع ما ، $\overline{P}$ تفسر على أنها نسبة الأفراد الذين لهم هذه الصفة في عينة أخذت من ذلك المجتمع.

مثال (15) :

إذا كان احتمال نجاح الطالب الذي يدرس أحد مقررات الاقتصاد هو 0.9 ، أخذت عينة حجمها 49 طالبا من أولئك الذي يدرسون هذا المقرر، أوجد $P(\overline{P} \geq 0.8)$.

الحل : من المعلوم أن $E[\overline{P}] = p = 0.9$

وتباين $\overline{P}$ هو $\frac{p(1-p)}{n} = \frac{0.9 \times 0.1}{49} = 0.0018$

وباستعمال نظرية (6) والتحويل إلى الطبيعي المعياري نجد :

$$P(\overline{P} \geq 0.8) = P\left(\frac{P-0.9}{\sqrt{0.0018}} \geq \frac{0.8-0.9}{\sqrt{0.0018}}\right)$$
$$= P\left(Z \geq \frac{-0.1}{0.042}\right)$$
$$= P(Z \geq -2.38)$$
$$= 1 - 0.0087 = 0.9918$$

مثال (16) :

أخذت عينة عشوائية حجمها 400 من مجتمع نسبة النجاح فيه p = 0.6. إذا كان $\overline{P}$ نسبة النجاح في العينة، أوجد $P(0.5 \leq \overline{P} \leq 0.72)$.

الحل : بما أن حجم العينة كبير فإنه يمكن تقريب توزيع $Z = \dfrac{\overline{P} - P}{\sqrt{\dfrac{P(1-P)}{n}}}$ بالتوزيع الطبيعي المعياري وعليه يكون :

$$P(0.5 \leq \overline{P} \leq 0.72) = P\left(\frac{0.5-0.6}{\sqrt{\dfrac{0.6 \times 0.4}{400}}} \leq Z \leq \frac{0.72-0.6}{\sqrt{\dfrac{0.6 \times 0.4}{400}}}\right)$$
$$= P\left(\frac{-0.1}{0.024} \leq Z \leq \frac{0.12}{0.024}\right)$$
$$= P(-0.41 \leq Z \leq 0.5)$$
$$= 0.6915 - 0.3409$$
$$= 0.3506$$

مثال (17) :

صف توزيع $\overline{P}$ للعينات ذات الحجم 100 من مجتمع فيه p= 0.4 . ما قيمة التوقع والخطأ المعياري في هذا التوزيع ؟

الحل : من نظرية (6) فإن توقع $\overline{P}$ هو :

$$E(\overline{P}) = p = 0.4$$

وتباين $\overline{P}$ هو : $\sigma_{\overline{P}}^2 = \dfrac{p(1-p)}{n}$

$$= \frac{0.4 \times 0.6}{100} = 0.0024$$

والخطأ المعياري الذي هو الانحراف المعياري للنسبة $\overline{P}$ يكون $\sqrt{0.0024} \cong 0.05$. أما توزيع $\overline{P}$ فيكون تقريبا التوزيع الطبيعي ذا الوسط 0.4 والانحراف المعياري 0.05.

أما توزيع الفرق بين نسبتين فيعطى بالنظرية التالية :

نظرية (7) :

إذا أخذت عينتان عشوائيتان حجماهما n_2, n_1 من مجتمعين مستقلين يخضع الأول لتوزيع بيرنوللي [ذي الحدين b(1،p)] والثاني يخضع لتوزيع ذي الحدين $b(1, p_2)$ فإن الفرق بين النسبتين من العينتين $\overline{P}_1 - \overline{P}_2$ يخضع تقريبا لتوزيع طبيعي وسطه $p_1 - p_2$ وتباينه $\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}$ عندما تكون n_2, n_1 كبيرتين.

مثال (18) :

إذا كانت نسبة النجاح في امتحان التوجيهي في مدرسة للبنات (p) هي 0.7 وكانت نسبة النجاح في ذلك الامتحان في مدرسة للذكور (ب) هي 0.65. أخذت عينة عشوائية حجمها $n_1 = 70$ من المدرسة (أ) وأخذت عينة عشوائية حجمها $n_2 = 35$ من المدرسة (ب) فما احتمال أن تزيد نسبة النجاح في المدرسة (أ) على نسبة النجاح في المدرسة (ب) بمقدار 0.10 على الأكثر ؟

الحل :

المطلوب إيجاد $P(\overline{P}_1 - \overline{P}_2 \leq 0.10)$

وبتطبيق النظرية (7)

$$P\left(\bar{P}_1-\bar{P}_2\le 0.10\right)=P\left(\frac{\bar{p}_1-\bar{p}_2-(P_1-P_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}}\le\frac{0.10-(0.70-0.65)}{\sqrt{\frac{0.7\times 0.30}{70}+\frac{0.65\times 0.35}{35}}}\right)$$

$$=P\left(Z\le\frac{0.05}{0.097}\right)=P(Z\le 0.51)$$

$$=0.6950$$

7 : 9 توزيع المعاينة للتباين وللنسبة بين تبايني عينتين

Sampling distribution of S^2 and S_1^2/S_2^2

نحتاج في كثير من الأحيان في تطبيقات الإحصاء الاستقرائي لمعرفة توزيع تباين العينة، وسنعطي هذا التوزيع في حالة ا لمعاينة من توزيع طبيعي.

نظرية (8) :

إذا أخذت عينة عشوائية حجمها n من توزيع طبيعي معدله μ وتباينه σ^2 أي $N(\mu,\sigma^2)$. وكان S^2 هو تباين العينة فإن :

$\frac{(n-1)S^2}{\sigma^2}$ يخضع لتوزيع كاي تربيع بدرجات حرية (n-1).

مثال (19) :

أخذت عينة عشوائية حجمها n = 15 من توزيع طبيعي $N(\mu,49)$ ، وكان S^2 تباين العينة فأوجد احتمال أن تكون S^2 أقل من 82.9.

(لاحظ أننا نسأل مثل هذا السؤال قبل معرفة قيم أفراد العينة وقبل حساب قيمة S^2).

الحل : المطلوب إيجاد $P(S^2\le 82.9)$.

باستعمال النظرية (8) :

$$P\left(S^2 \le 82.9\right) = P\left(\frac{(n-1)S^2}{\sigma^2} \le \frac{14 \times 82.9}{49}\right)$$
$$= P\left(\chi^2 \le 23.685\right) = 0.95$$

حيث χ^2 يخضع لدرجات حرية 14

للمقارنة بين تبايني مجتمعين فإننا نحتاج النسبة بين تبايني عينتين مأخوذتين من هذين المجتمعين وسنعطي توزيع هذه النسبة في حالة المعاينة من مجتمعين طبيعيين مستقلين.

نظرية (9) : ليكن S_1^2 تباين عينة عشوائية حجمها n_1 من مجتمع $N\left(\mu_1, \sigma_1^2\right)$

وليكن S_2^2 تباين عينة عشوائية حجمها n_2 من مجتمع $N\left(\mu_2, \sigma_2^2\right)$ مستقل عن الأول ، فإن :

$$\frac{S_1^2 / S_2^2}{\sigma_1^2 / \sigma_2^2} = \frac{S_1^2}{S_2^2} \frac{\sigma_2^2}{\sigma_1^2}$$

يخضع لتوزيع F ذي درجات الحرية $\left(n_1 - 1, n_2 - 1\right)$.

مثال (20) : أخذت عينة عشوائية حجمها 11 من توزيع $N\left(\mu_1, \sigma^2\right)$ وأخذت عينة عشوائية حجمها 16 من توزيع $N\left(\mu_2, \sigma^2\right)$ مستقل عن الأول . أوجد :

$$P\left(\frac{S_1^2}{S_2^2} \ge 3.80\right)$$

وذلك قبل معرفة قيمة S_1^2 / S_2^2.

الحل بما أن تباين كل من المجتمعين هو σ^2 (أي تباينا المجتمعين متساويان) فإن توزيع $\frac{S_1^2}{S_2^2}$ هو توزيع F بدرجات الحرية (15, 10).

ومن جدول توزيع F بدرجات الحرية (15, 10) نجد :

$$P\left(\frac{S_1^2}{S_2^2} \geq 3.80\right) = 1 - 0.99 = 0.01$$

تمارين

1-7 : إذا أردت أخذ عينة حجمها 300 من طلبة احدى الجامعات ذات الكليات: العلوم، الهندسة، الصيدلة، والآداب والتي عدد الطلبة فيها على التوالي 600، 800، 400، 1200 فما عدد أفراد العينة من كل كلية إذا أردت أن تكون جميع الكليات ممثلة بنفس النسبة.

2-7 : مجتمع فيه 620 عنصرا، اختر عينة عشوائية بسيطة حجمها 10 باستعمال جدول الأعداد العشوائية.

3-7 : يرغب مدير بنك في دراسة 1000 قرضا منحها البنك في العام السابق. من الأسئلة التي يرغب الإجابة عليها، ما نسبة القروض التي تزيد قيمته عن 10000 دينار أردني ؟ ما معدل قيمة القروض التي منحها البنك ؟.

إن الإجابات عن هذه الأسئلة تمكن مدير البنك من تقدير الطلب على المبالغ النقدية في المستقبل.

(أ) عين المجتمع تحت الدراسة.

(ب) عين المتغيرات تحت الاهتمام. هل هذه المتغيرات متصلة أم منفصلة ؟

(جـ) عين معلمات المجتمع الهامة.

(د) هل المجتمع موجود أم أنه فرضي ؟ وضح.

(هـ) أخذت عينة حجمها 100 قرض ووجد أن 40 منها تزيد عن 10000دينار أردني. هل تستنتج أن نسبة القروض التي تزيد قيمتها عن 10000 دينار هي بالضبط 0.40 ؟ وضح.

4-7 : تخضع معاملات الذكاء لطلبة الصفوف الأساسية في احدى المحافظات لتوزيع وسطه 107 وتباينه 64.

أخذت عينة عشوائية حجمها 100 من هؤلاء الطلبة، فما احتمال أن يقع الوسط الحسابي لمعاملات الذكاء في العينة ما بين 105، 110 متضمنة ؟

5-7 : يخضع طول الطالبات في المدارس الإعدادية لتوزيع وسطه 152سم وانحرافه المعياري 6 سم فما احتمال أن يزيد الوسط الحسابي لأطوال عينة مؤلفة من 64 طالبة على 153 سم؟

وما احتمال أن ينقص الوسط الحسابي لأطوال تلك العينة عن 150سم ؟

7-6 : تخضع أوزان الأطفال عند الولادة في بلد ما ، لتوزيع غير ملتو كثيرا وسطه 3.3 كغم وانحرافه المعياري 1.1 كغم إذا اخترنا عينة عشوائية حجمها 40 طفلا، فما احتمال أن يزيد معدل الأوزان لهذه العينة على 3.6 كغم ؟

وما احتمال أن ذلك المعدل ينقص عن 2.8 كغم ؟

7-7 : إذا كان الزمن الذي يستغرقه محاسب الصندوق لإنجاز معاملة الزبون في مصرف ما يخضع لتوزيع وسطه 75 ثانية وتباينه 100 ثانية تربيع. ما احتمال أن معدل الزمن الذي يستغرقه 50 محاسبا لإنجاز مثل هذه المعاملات يقل عن 72 ثانية.

ما احتمال أن ذلك المعدل يزيد على 80 ثانية ؟

7-8 : يخضع طول الطلاب في المرحلة الاعدادية في مدينة ما لتوزيع طبيعي وسطه 158 سم وانحرافه المعياري 6 سم.

ما احتمال أن يقع الوسط الحسابي لأطوال 16 طالباً ما بين 153، 160 ؟

7-9 : يخضع طول الحياة لنوع من إطارات السيارات لتوزيع طبيعي وسطه 23000 كم وانحرافه المعياري 2000 كم، فإذا أخذنا منها 4 إطارات، فما احتمال أن يزيد معدل حياتها على 23800 كم ؟

7-10 : أخذت عينة عشوائية حجمها 400 من توزيع نسبة النجاح فيه 0.4 أوجد :

(أ) $\mathbf{P(0.37 < \overline{P} < 0.43)}$

(ب) $\mathbf{P(\overline{P} < 0.36)}$

حيث $\overline{\mathbf{P}}$ هي نسبة النجاح في العينة.

7-11 : لوحظ أن نسبة المتخلفين عن المقابلة ممن يتقدمون لشغل وظائف في وزارة التربية والتعليم عادة ما تكون 0.1. إذا أخذنا عينة عشوائية حجمها 100 من المطلوبين للمقابلة فما احتمال أن تقل نسبة عدد المتخلفين عن 0.15 ؟

7-12 : على فرض أن الرأي العام في بلد كان منقسما إلى مؤيدين : غير مؤيدين بنسبة هي 15:85 حول موضوع التأمين الصحي. أخذت آراء عينة من 625 شخصا فما احتمال أن تزيد نسبة المؤيدين للتأمين الصحي على 90% ؟

7-13 : أخذت عينة عشوائية حجمها 30 من مجتمع بيرنوللي فيه نسبة النجاح 0.7 وأخذت عينة عشوائية حجمها 45 من مجتمع آخر مستقل عن الأول ونسبة النجاح فيه 0.55.

فإذا كانت نسبة النجاح في العينة الأولى $\overline{\mathbf{P}}_1$ وفي العينة الثانية $\overline{\mathbf{P}}_2$ ، أوجد $\mathbf{P}\left(\overline{\mathbf{P}}_1 - \overline{\mathbf{P}}_2 \geq \mathbf{0.2}\right)$.

7-14 : تخضع علامات الناجحين في امتحان الدراسة الثانوية العامة في إحدى المدارس (أ) لتوزيع طبيعي معدله 70 وانحرافه المعياري 12، وفي مدرسة ثانية (ب) تخضع العلامات لتوزيع طبيعي معدله 76 وانحرافه المعياري 16.

أخذت عينة عشوائية حجمها 40 من المدرسة (أ) وعينة عشوائية حجمها 20 من المدرسة (ب). على فرض أن الوسط الحسابي للعينة الأولى $\overline{\mathbf{X}}$ وللعينة الثانية $\overline{\mathbf{Y}}$ أوجد :

(أ) $\mathbf{P}\left(\overline{\mathbf{Y}} - \overline{\mathbf{X}} \geq \mathbf{8}\right)$

(ب) $\mathbf{P}\left(\left|\overline{\mathbf{Y}} - \overline{\mathbf{X}}\right| \leq \mathbf{3}\right)$

7-15 : أخذت عينة عشوائية حجمها 10 من مجتمع طبيعي تباينه 8، أوجد الثابت C بحيث :

$$P\ (S^2 \leq C) = 0.95$$

حيث S^2 هو تباين العينة.

7-16 : أخذت عينة عشوائية حجمها 9 من مجتمع طبيعي تباينه 16 وأخذت عينة عشوائية ثانية حجمها 15 من مجتمع طبيعي مستقل عن الأول وتباينه 12. فإذا كان $\mathbf{S}_2^2,\mathbf{S}_1^2$ هما تبايني العينتين الأولى والثانية ، أوجد الثابت C بحيث :

$$\mathbf{P}\left(\frac{\mathbf{S}_1^2}{\mathbf{S}_2^2} \leq \mathbf{C}\right) = \mathbf{0.90}$$

أوجد الثابت a بحيث :

$$\mathbf{P}\left(\frac{\mathbf{S}_2^2}{\mathbf{S}_1^2} \leq \mathbf{a}\right) = \mathbf{0.95}$$

أوجد الثابت b بحيث :

$$P\left(\frac{S_1^2}{S_2^2} \geq b\right) = 0.05$$

الفصل الثامن
التقدير
Estimation

8-1 مقدمة :

الاستنتاجات الإحصائية هي التعميمات والقرارات التي يمكنك اتخاذها بناء على معلومات أو بيانات قمت بجمعها أو كانت متوفرة لديك، فمثلا إذا دخلت حقلا فيه عدد كبير من أشجار الزيتون أو البرتقال وقمت بتفحص عدد من هذه الأشجار ولاحظت أنها ذات ثمار جمة، فإنك ولا شك ستستنتج أن محصول هذه الأشجار جيد وأنها عالية الإنتاج في ذلك العام. إن هذا يعني أنك بنيت رأيك عن كمية المحصول المنتظر بدراسة محصول عينة من الأشجار التي قمت بملاحظتها. وكمثال آخر، إذا استعملت عددا من المصابيح الكهربائية من إنتاج مصنع ما ووجدت أن حياة هذه المصابيح قصيرة، ولنقل حوالي 600 ساعة، فهل تستمر في استعمال هذا النوع من المصابيح إذا علمت أن عينة من مصابيح مصنع آخر أعطت معدل حياة 1000 ساعة ؟

هنا أيضا ستقرر قرارا مبنيا على مقارنة بين نتائج عينة المصابيح التي استعملتها ونتائج العينة الثانية.

إذا أرادت شركة أدوية أن تسوق دواء فإنه يجب أن تحصل على تصريح باستعمال ذلك الدواء أولا، وإحدى الطرق التي تتقدم بها الشركة للحصول على التصريح هي تقديم إثبات أن الدواء قد جرب فثبتت جدوى استعماله. وهذا يعني أن عينة من المرضى قد استعملوا الدواء وأفادهم. وأن الشركة بنت قرارها من دراسة تلك العينة.

في كل من الأمثلة السابقة قمنا بدراسة عينة وبناء على نتائجها تقرر تعميمها على مجتمع، واعتبرنا أن ما حصلنا عليه من العينة من النتائج يصلح للمجتمع.

أما إلى أي مدى ستكون التعميمات صحيحة فهو أمر يتطلب الإجابة عنه دراسة الاحتمالات وتوزيعات المعاينة واستعماله بشكل منطقي يحدد مدى دقة تمثيل العينة للمجتمع ، ومدى دقة تصميم نتائج العينة عليه أيضا.

من الأمثلة السابقة، نستخلص أن من أهم فروع الإحصاء الاستنتاجي : التقدير واختبار الفرضيات وسندرس في هذا الفصل التقدير وندرس اختبار الفرضيات في الفصل التاسع.

8 : 2 التقدير Estimation

يتم التقدير باختيار عينة من المجتمع ومشاهدة مفردات تلك العينة، ومن ثم حساب المقياس المراد وتعميم ذلك على المجتمع. فلو أردت تقدير ناتج حقل من الزيتون تأخذ عينة عشوائية من عشرة زيتونات على سبيل المثال وتزن إنتاجها جميعها ثم تقسم على 10 فتحصل على معدل إنتاج الشجرة الواحدة في العينة ونستعمل هذا المعدل لتقدير معدل إنتاج الشجرة في الحقل بأكمله.

وإذا علمت عدد الأشجار في الحقل فإنك تقدر الإنتاج الكامل على أنه يساوي عدد الأشجار مضروبا في معدل إنتاج الشجرة الواحدة.

للتقدير أهمية كبرى وله مجالات تطبيقية في الزراعة والصناعة والتربية والدراسات الاجتماعية والصحية. فلو أرادت وزارة الصحة تقدير نسبة الطلبة في المدارس الأساسية الذين يحتاجون إلى نظارات طبية، فكيف ستقوم بذلك؟ إذا أمعنت النظر فلابد أنك ستجيب بأن على الوزارة دراسة حالة عينة ممثلة للطلبة المذكورين ومعرفة نسبة المحتاجين إلى استعمال النظارة الطبية، ومن ثم تعميم ذلك على مجتمع الطلبة، بذلك تكون قد

أعطيت تقديرا للنسبة في المجتمع من دراستك للنسبة في العينة.

هذا النوع من التقدير يسمى التقدير بنقطة، ولما كان من غير المحتمل معرفة النسبة الحقيقية (النسبة في المجتمع) بالضبط فإنك تحتاج إلى إعطاء فترة يكون من الممكن وقوع نسبة المجتمع داخل هذه الفترة. أي أنك تعطي قيمتين وتقول إن نسبة المجتمع تقع بين هاتين القيمتين،، ولما كان لابد من وجود خطأ في تقديرك، فإنك تقول إن نسبة المجتمع تقع بين قيمتين باحتمال معين. وهذا النوع من التقدير يسمى التقدير بفترة. وكما أنك تحتاج إلى تقدير النسبة فإنك تحتاج أيضا إلى تقدير الوسط الحسابي، وتقدير التباين ، وغيرها وأهم أنواع التقدير التي ندرسها هي التقدير النقطي والتقدير بفترة.

8 : 3 التقدير النقطي Point Estimation

التقدير النقطي هو تقدير إحدى معلمات المجتمع بنقطة وذلك بإعطاء قيمة واحدة لتلك المعلمة. فلو أردت أن تقدر معدل طول حياة المصابيح الكهربائية المنتجة في مصنع معين فماذا تعمل ؟ إن إحدى الطرق التي ستقوم بها هي أن تأخذ عينة عشوائية من إنتاج هذا المصنع، وتجري عليها الاختبار ، وتسجل طول حياة كل مصباح منها، ثم تحسب الوسط الحسابي لأطوال الحياة. إن القيمة التي تحصل عليها هي وسط العينة وهي القيمة التي تستعملها كتقدير لوسط المجتمع.

وكما درسنا سابقا فإن :

تعريف (1)

المعلمة Parameter هي ثابت يصف المجتمع أو يصف توزيع المجتمع كالوسط الحسابي للتوزيع أو الانحراف المعياري له.

تعريف (2)

الإحصاء Statistic هو اقتران تتعين قيمته من العينة، كالوسط الحسابي للعينة أو لانحراف المعياري لها.

8 : 4 التقدير النقطي بطريقة العزوم

Point estimation by the method of moments

هناك طرق عديدة لإيجاد التقدير النقطي لمعلمة مجتمع ما ، وإن من أبسطها طريقة العزوم.

وتتلخص هذه الطريقة بأنه إذا كان للمجتمع معلمة مجهولة واحدة فإننا نساوي العزم الأول للمجتمع بالعزم الأول للعينة ونحل المعادلة الناتجة لإيجاد تقدير المعلمة بدلالة العزم الأول للعينة.

مثال (1) :

إذا أخذت عينة عشوائية حجمها n من توزيع بواسون $P(\lambda)$ وإذا عبرنا عن أفراد العينة بالحروف $X_1, X_2, \dots, X_n$ فمن المعلوم أن العزم الأول للمجتمع هو $E[X] = \lambda$ وأن العزم الأول للعينة هو $\overline{X} = \frac{\sum X_i}{n}$ وبوضع $E(X) = \overline{X}$ نجد أن $\hat{\lambda} = \overline{X}$ أي أن تقدير λ بطريقة العزوم هو $\hat{\lambda} = \overline{X}$.

فإذا كانت قيم العينة من توزيع بواسون $P(\lambda)$ هي 2 ,5 ,3 ,0 ,4 ,9 ,5 فإن التقدير النقطي للمعلمة λ هو :

$$\overline{X} = \frac{2+5+3+0+4+9+5}{7} = 4$$

مثال (2) :

إذا أخذت عينة عشوائية حجمها n معلوم من مجتمع بيرنوللي أي ذات الحدين $b(1, p)$ فما هو التقدير النقطي للمعلمة p.

نتذكر أولا إذا عبرنا عن متغير بيرنوللي بالرمز X فهذا يعني أن $P(X = 0)=1-p$, $P(X = 1) = p$ وبذلك فإن $E[X] = p$.

وإذا عبرنا عن أفراد العينة بالرموز :

$X_1, X_2, \dots, X_n$

فإن العزم الأول للعينة هو $\overline{X} = \frac{\sum X_i}{n}$ وبالمساواة بين الطرفين نجد $\hat{P} = \overline{X}$ هو التقدير النقطي للمعلمة p.

وبما أن X_i لجميع $i = 1, 2, \dots, n$ تأخذ القيم 1 , 0 فقط (نجاح ، فشل) أو (نعم , لا) ، وهكذا ، فإن $\overline{X}$ هي نسبة النجاح في العينة ونعبر عنها بالرمز $\overline{P}$.

فمثلا، إذا كانت نتيجة عينة من الطلبة في امتحان التوجيهي اخترناها عشوائيا من احدى المدارس كما يلي :

ناجح، ناجح،ناجح، راسب، ناجح، راسب، راسب ناجح، ناجح، راسب. فما هو تقدير نسبة النجاح في المدرسة ؟

إن نسبة النجاح في العينة هي $\bar{p} = \frac{6}{10}$ إذاً نقدر نسبة النجاح في المدرسة بقيمة $\bar{P}$ وهي 0.6.

أما إذا كان في المجتمع معلمتان مجهولتان فإننا نجد تقديريهما بطريقة العزوم بمساواة العزم الأول للمجتمع بالعزم الأول للعينة ومساواة العزم الثاني للمجتمع بالعزم الثاني للعينة ومن ثم حل هاتين المعادلتين لنجد تقدير المعلمتين بدلالة العزمين الأول والثاني للعينة.

مثال (3) :

إذا كانت $X_1, X_2, \ldots, X_n$ عينة عشوائية من مجتمع طبيعي $N(\mu, \sigma^2)$ ، فلإيجاد التقدير النقطي بطريقة العزوم لكل من μ, σ^2 نجد العزمين الأول والثاني للعينة وهما $\bar{X}$, $\sum_{i=1}^{n} X_i^2 / n$ ، ونجد العزمين الأول والثاني للمجتمع وهما $E(X^2), \mu$ ونجد $E(X^2)$ من المعادلة :

$$\sigma^2 = E(X^2) - \mu^2$$

إذاً

$$E(X^2) = \sigma^2 + \mu^2$$

الآن نضع المساواة بين كل عزمين من نفس الرتبة، أي $\mu = \bar{X}$

$$\frac{\sum X_i^2}{n} = \sigma^2 + \mu^2$$

ويحل هاتين المعادلتين نجد : $\hat{\mu} = \bar{X}$

$$\hat{\sigma}^2 = \frac{\sum X_i^2}{n} - (\bar{X})^2$$

وهما التقديران النقطيان بطريقة العزوم للمعلمتين μ, σ^2 على التوالي.

5 : 8 خصائص المقدر الجيد Properties of Good Estimator :

إننا نقدر معلمة المجتمع بمقدر معين، وهو إحصاء، لذلك فهو متغير عشوائي تتعين قيمته من العينة تحت الدراسة، فمثلا إذا كان وسط المجتمع μ نأخذ عينة عشوائية حجمها n ونشاهد أفراد العينة ونحسب قيمة وسطها الحسابي $\bar{X}$.

وإذا أخذنا عينة ثانية فربما نحصل على وسط حسابي آخر. ويمكن أن نقدر وسط المجتمع μ بقيمة مشاهدة واحدة من المجتمع ولتكن X_1 أو أن نستعمل قيمتين ونأخذ وسطهما الحسابي $\frac{X_1 + X_2}{2}$ ونهمل بقية المشاهدات ولكن ما يتفق مع المنطق فهو أن نستعمل الوسط الحسابي $\overline{X}$ كتقدير للمعلمة μ.

بما أن المقدر متغير عشوائي كان منطقيا أن نعتبر المقدر الجيد هو ذلك المقدر الذي يتمركز حول المعلمة. أي أنه من الممكن أن نعتبر المقدر جيدا إذا كان ذلك المقدر قريبا من المعلمة بالمعدل، وهذا يقودنا إلى:

تعريف (3) :

إذا كانت معلمة المجتمع θ واستعملت المقدر $\hat{\theta}$ لتقدير قيمتها، فإنك تقول إن $\hat{\theta}$ مقدر غير متحيز إذا كان توقع $\hat{\theta}$ يساوي θ. أي أن $E(\hat{\theta}) = \theta$.

إذا أخذت عينة عشوائية من مجتمع وسطه μ وتباينه σ^2 وكان حجم العينة n، هل X_1 غير متحيز ؟

الجواب نعم لأن $E(X_1) = \mu$.

هل $\frac{X_1 + X_2}{2}$ غير متحيز ؟

الجواب نعم ، لأن.

$$E\left(\frac{X_1 + X_2}{2}\right) = E\left(\frac{X_1}{2}\right) + E\left(\frac{X_2}{2}\right)$$
$$= \frac{1}{2}\mu + \frac{1}{2}\mu = \mu$$

وبنفس الطريقة تجد أن $\overline{X}$ غير متحيز.

وعليه، عرفنا أن X_1 , $\frac{X_1 + X_2}{2}$, $\overline{X}$ كلها مقدرات غير متحيزة للمعلمة μ.

الآن، هل يوجد فروق بين هذه المقدرات الثلاثة ؟ أي هذه المقدرات أفضل من غيره ؟

احسب التباين لكل من المقدرات آنفة الذكر :

(1) تباين X_1 هو تباين المجتمع ويساوي σ^2 .

(2) تباين $\frac{X_1+X_2}{2}$ هو (استعمل كلمة Var لتعني التباين).

$$Var\left(\frac{X_1+X_2}{2}\right) = Var\left(\frac{X_1}{2}\right) + Var\left(\frac{X_2}{2}\right)$$
$$= \frac{1}{4}\sigma^2 + \frac{1}{4}\sigma^2 = \frac{\sigma^2}{2}$$

(3) تباين $\overline{X}$ هو $Var(\overline{X}) = \sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$

بمقارنة (1)، (2)، (3) تجد أن تباين $\overline{X}$ هو أصغر التباينات الثلاثة عندما تكون $n > 2$، بالطبع إذا كان $n = 2$ فإن $\overline{X} = \frac{X_1+X_2}{2}$ بما أن تباين $\overline{X}$ هو الأصغر من بين الحالات الثلاث فهذا يعني أن تباين توزيع $\overline{X}$ أصغر من تباين توزيع X_1 أو $\frac{X_1+X_2}{2}$ وهذا يعني تباعد القيم $\overline{X}$ عن μ أقل من تباعد القيم X_1 أو القيم $\frac{X_1+X_2}{2}$ عن μ نفسها عندما $n > 2$.

إذن من المنطقي أن نختار المقدر الذي تباينه أصغر من تباينات المقدّرات الأخرى.

ومن هنا ، إذا كان $\hat{\theta}_2, \hat{\theta}_1$ تقديرين غير متحيزين للمعلمة θ في مجتمع ما فإنك تقول إن $\hat{\theta}_1$ أكثر كفاية من $\hat{\theta}_2$ إذا كان $\sigma^2_{\hat{\theta}_1} < \sigma^2_{\hat{\theta}_2}$.

تعريف (5) :
إذا اعتبرت جميع المقدرات غير المتحيزة θ في مجتمع ما ، فإن المقدر ذا أصغر تباين يكون المقدر الأكثر كفاية (الأفضل) لتقدير المعلمة θ.

مثال (4) :

أخذت عينة عشوائية حجمها 9 من مجتمع طبيعي وسطه μ وتباينه σ^2 فأي المقدرات (لتقدير μ) التالية غير متحيز. أي منها هو الأفضل ؟

$$X_5, \bar{X}, \frac{X_2 + X_3}{2}, X_1$$

الحل : العينة العشوائية هي $X_1, X_2, \ldots, X_9$ من الواضح أن

$$E(X_5) = \mu, E(\bar{X}) = \mu, E\left(\frac{X_2 + X_3}{2}\right) = \mu, E(X_1) = \mu$$

إذن جميع المقدرات المذكورة غير متحيزة.

تقارن الآن بين تباينات هذه المقدرات

1) $Var(X_1) = \sigma^2$.

2) $Var\left(\frac{X_2 + X_3}{2}\right) = \frac{1}{4}\sigma^2 + \frac{1}{4}\sigma^2 = \frac{\sigma^2}{2}$

3) $Var(\bar{X}) = \sigma^2_{\bar{X}} = \frac{\sigma^2}{9}$

4) $Var(X_5) = \sigma^2$

ستجد أن أفضل المقدرات المذكورة هو $\bar{X}$ لأن تباينه أصغر تباين بينها.

8 : 6 التقدير بفترة (فترات الثقة)

Interval Estimation (Confidence Intervals)

لقد ذكرنا سابقا أننا لا نتوقع الحصول على تقدير لمعلمة المجتمع بدون خطأ مهما كان هذا التقدير جيدا، وبعبارة أخرى ليس من المحتمل أن تحصل على تقدير يقدّر معلمة المجتمع تماما. ومع أن دقة التقدير تزداد بزيادة حجم العينة فإنه ليس هناك أي سبب يبرر إمكانية الحصول على تقدير يقوم بتقدير معلمة المجتمع بدون خطأ، أي أن قيمته تساوي قيمة معلمة المجتمع بالضبط، ولذلك فإنه من المرغوب فيه إعطاء فترة معينة تتوقع أن تقع معلمة المجتمع داخلها. إن مثل هذه الفترة تسمى فترة تقدير أو فترة ثقة.

ولتوضيح هذا المفهوم، نفرض أننا أخذنا عينة عشوائية حجمها n من مجتمع طبيعي $N(\mu,\sigma^2)$ حيث σ^2 معلومة. افرض أن الوسط الحسابي للعينة هو $\overline{X}$. والمطلوب إيجاد فترة ثقة للمعلمة μ.

إننا نعلم أن $\overline{X}$ متغير عشوائي توزيعه التوزيع الطبيعي ذو الوسط μ والتباين $\frac{\sigma^2}{n}$ أي أن $Z=\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ يخضع للتوزيع الطبيعي المعياري $N(0,1)$.

ومن جداول التوزيع الطبيعي المعياري نجد :

$$P(-1.96\le Z\le 1.96)=0.95$$

إذن

$$P\left(-1.96\le \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\le 1.96\right)=0.95$$

وبالتبسيط الجبري (وذلك بالضرب بـ $\frac{\sigma}{\sqrt{n}}$ ثم طرح $\overline{X}$ ثم الضرب بـ 1- مع تغيير ≤ إلى ≥) نجد :

$$P\left(\overline{X}-1.96\frac{\sigma}{\sqrt{n}}\le \mu\le \overline{X}+1.96\frac{\sigma}{\sqrt{n}}\right)=0.95$$

إن هذه المعادلة تعني انه قبل إجراء المشاهدات على العينة وقبل حساب قيمة $\overline{X}$ من العينة فإن الفترة التي حدّها الأيسر $\overline{X}-1.96\frac{\sigma}{\sqrt{n}}$ (وهو متغير عشوائي) وحدّها الأيمن $\overline{X}+1.96\frac{\sigma}{\sqrt{n}}$ (وهو متغير عشوائي أيضا). تحوي المعلمة المجهولة القيمة μ باحتمال 0.95.

ومن هذا المنطلق نقول إن الفترة :

$$\left(\overline{X}-1.96\frac{\sigma}{\sqrt{n}},\overline{X}+1.96\frac{\sigma}{\sqrt{n}}\right)$$

هي فترة ثقة 95% للمعلمة μ.

وكذلك ، وبعد حساب الوسط الحسابي وإيجاد قيمته ولتكن $\overline{x}$ ، فإننا نقول إن الفترة $\left(\overline{x}-1.96\frac{\sigma}{\sqrt{n}},\overline{x}+1.96\frac{\sigma}{\sqrt{n}}\right)$ هي فترة ثقة 95% للمعلمة μ.

مثال (5) :

أخذت عينة عشوائية حجمها n =25 من مجتمع $N(\mu,16)$ فوجد أن $\overline{X}=12.5$ أوجد فترة ثقة 95% للمعلمة المجهولة μ.

الحل : المجتمع طبيعي وتباينه معلوم وقيمة الوسط الحسابي $\overline{X}=12.5$.

إذن فترة ثقة 95% هي :

$$\left(12.5-1.96\times\frac{4}{5},12.5+1.96\times\frac{4}{5}\right)$$

أي (10.93 , 14.068) هي فترة ثقة 95% للمعلمة μ.

لاحظ أن العدد 1.96 حصلنا عليه من التوزيع الطبيعي المعياري جوابا للسؤال : ما هما النقطتان المتماثلتان اللتان تحصران بينهما 0.95 من المساحة تحت التوزيع الطبيعي المعياري، وكان الجواب 1.96 +، 1.96-.

ولذلك لو أردنا فترة ثقة 90% للمعلمة μ تحت الشروط السابقة لاستعملنا القيمتين1.645,1.645-ولو أردنا فترة ثقة 99% فإننا نستعمل القيمتين2.33, 2.33-.

وبشكل عام ، لو أردنا فترة ثقة $100(1-\alpha)\%$ للمعلمة μ تحت الشروط السابقة فإننا نستعمل القيمتين $-z_{1-\frac{\alpha}{2}}, z_{1-\frac{\alpha}{2}}$ حيث أنهما النقطتان المتماثلتان اللتان تحصران بينهما $(1-\alpha)$ من المساحة تحت التوزيع الطبيعي المعياري، وبهذا نحصل على :

نظرية (1) : إذا أخذت عينة عشوائية حجمها n من مجتمع $N(\mu,\sigma^2)$ وكانت σ^2 معلومة فإن فترة ثقة $100\%(1-\alpha)$ للمعلمة μ هي :

$$\left(\overline{X}-z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \overline{X}+z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$

حيث $\overline{X}$ هو الوسط الحسابي للعينة ، $z_{1-\frac{\alpha}{2}}$ هي القيمة على محور z التي إلى يسارها مساحة $1-\frac{\alpha}{2}$.

تفسير فترة الثقة :

تعتبر فترة الثقة من الأدوات القوية التي تعطي معلومات عن المعلمة المجهولة (مثل μ) باستعمال العينة ، ولذلك نعطي بعض الملاحظات حول طبيعة فترة الثقة، ونكتفي بالشرح عن فترة ثقة 95% لأن فترة ثقة $100\%(1-\alpha)$ لها نفس السلوك.

(1) قبل دراسة العينة وتسجيل المشاهدات وإيجاد قيمة الوسط الحسابي فإن :

$$\left(\overline{X}-1.96\frac{\sigma}{\sqrt{n}}, \overline{X}+1.96\frac{\sigma}{\sqrt{n}}\right)$$

هي فترة نهايتاها متغيران عشوائيان أي أنها فترة عشوائية تحاول احتواء المعلمة المجهولة μ.

(2) إن تفسير الاحتمال

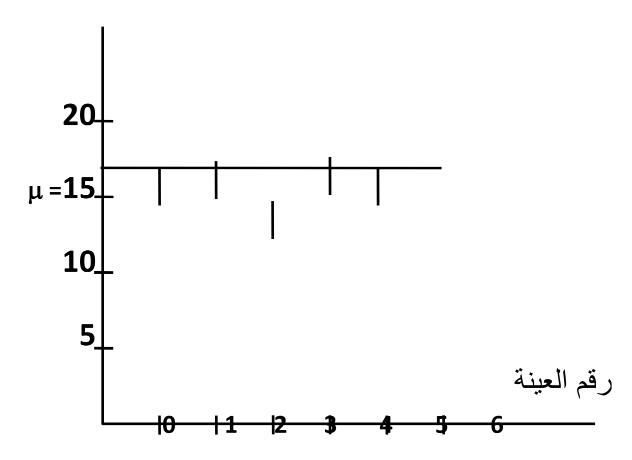
$$P\left(\overline{X}-1.96\frac{\sigma}{\sqrt{n}}<\mu<\overline{X}+1.96\frac{\sigma}{\sqrt{n}}\right)$$

على أنه التكرار النسبي لمحاولات المعاينة الكثيرة المتكررة يحدد أن 95% من فترات الثقة ستحوي μ وأن 5% منها لا تحويها.

(3) حالما نحسب الوسط الحسابي من العينة ونجد قيمته $\bar{x}$ فإن فترة الثقة $\left(\bar{x}-1.96\frac{\sigma}{\sqrt{n}}<\mu<\bar{x}+1.96\frac{\sigma}{\sqrt{n}}\right)$ تسمى فترة ثقة 95% للمعلمة μ، وهي فترة ثقة محددة عدديا ولا معنى للحديث عندها عن احتمال احتواء فترة معلومة قيمة ثابتة غير معلومة مثل μ، فهي إما تحويها أو لا تحويها.

(4) وعند التطبيق فنحن لا نعلم فيما إذا كانت فترة الثقة 95% التي حصلنا عليها من عينة معينة، تحوي القيمة المجهولة μ أو لا تحويها. ولكن اعتمادا على التكرار النسبي للمحاولات الكثيرة التي شرحناها في (2) فقد أمكننا استعمال كلمة ثقة بعد حساب الفترة.

وبأخذ هذه الملاحظات في عين الاعتبار فإن تفسير فترة ثقة 95% للمعلمة μ هو: إذا أخذت مائة عينة عشوائية ذات الحجم n وفي كل مرة نحسب $\bar{X}$ ونحسب فترة الثقة فإنا نتوقع أن 95 فترة تحوي μ الحقيقة.



لاحظ أن العينات (1) ، (2) ، (4) ، (5) تعطى فترات ثقة تحوي μ أما العينة (3) فإنها تعطي فترة ثقة لا تحوي μ.

مثال (6) :

أوجد فترة ثقة 95% للمعدل μ في مجتمع طبيعي تباينه 64 إذا اخترت عينة عشوائية حجمها 9 وكان وسطها الحسابي $\bar{X}=32$.

الحل : بما أن $1-\alpha=0.95$

إذن $\frac{\alpha}{2} = 0.025$ وبالتالي $z_{1-\alpha/2} = 1.96$

إذن ، بالتعويض في فترة الثقة تجد أن فترة 95% ثقة للوسط μ هي :

$$32 - 1.96 \times \frac{8}{\sqrt{9}} < \mu < 32 + 1.96 \times \frac{8}{\sqrt{9}}$$
$$32 - 5.227 < \mu < 32 + 5.227$$

أي أن : $26.773 < \mu < 37.227$

هي فترة ثقة 95% للوسط μ.

مثال (7) :

عينة عشوائية حجمها n = 16 أخذت من مجتمع طبيعي انحرافه المعياري σ = 4 أعطت المعدل $\overline{X} = 60$.

أوجد فترة ثقة 98% لوسط المجتمع μ.

الحل : بما أن 1 - α = 0.98

إذن $\frac{\alpha}{2} = 0.01$ وبالتالي $z_{1-\alpha/2} = 2.33$.

بالتعويض في فترة الثقة كما في النظرية (1).

تجد أن فترة الثقة 98% للوسط μ هي :

$$60 \pm 2.33 \times \frac{4}{\sqrt{16}}$$

وبالاختصار تجد :

$57.67 < \mu < 62.33$ هي فترة الثقة المطلوبة.

8 : 7 فترات الثقة في حالة حجم العينة كبير

أعطت النظرية (1) فترة الثقة للوسط μ إذا كانت المعاينة من مجتمع طبيعي ذي تباين معلوم σ^2 ، والآن كيف نبني فترة ثقة للوسط μ إذا لم يكن المجتمع خاضعا للتوزيع الطبيعي ؟ إن نظرية تقارب

التوزيعات Central Limit Theorem تعطينا حجر الأساس حيث تعطينا $Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$ يخضع لتوزيع طبيعي معياري تقريبا إذا كانت n كبيرة ولذلك نستطيع استعمال النظرية :

نظرية (2) : إذا أخذت عينة عشوائية حجمها n من مجتمع وسطه μ وتباينه σ^2 معلومة فإن فترة ثقة $(1-\alpha)$100% للمعلمة μ هي تقريبا :

$$\left(\overline{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$$

إذا كانت n كبيرة $(n \geq 30)$.

مثال (8) :

عينة عشوائية حجمها n = 81 من مجتمع تباينه σ^2=36 أعطت $\overline{X}$ =33 ، أوجد فترة ثقة 98% لوسط المجتمع μ.

الحل: $1-\alpha = 0.98$ إذن $\frac{\alpha}{2} = 0.01$ ،ومن جدول الطبيعي المعياري $z_{0.99} = 2.33$

وبالتعويض في فترة الثقة (نظرية(2) نجد :

$$33 - 2.33 \times \frac{6}{9} < \mu < 33 + 2.33 \times \frac{6}{9}$$

وبالاختصار نحصل على :

$$31.447 < \mu < 34.53$$

و هي فترة ثقة 98% للوسط μ.

في كثير من الأحيان لا تكون σ^2 معلومة ولذلك نستعمل S^2 (تباين العينة كتقدير لـ σ^2) وتتمكن من استعمال النظرية (2) بوضع S بدلا من σ إذا كان حجم العينة n كبيرا $(n \geq 30)$.

مثال (9) :
عينة عشوائية حجمها n = 81 أعطت s = 6 ، $\bar{x} = 73$.

أوجد فترة ثقة 98% لوسط المجتمع μ.

الحل : من الواضح أن تباين المجتمع σ^2 غير معلوم لكن حجم العينة أكبر من 30 ولذلك تستعمل S بدلا من σ.

من الواضح أن $1 - \alpha = 0.98$ ولذلك $\frac{\alpha}{2} = 0.01$ ومن جدول الطبيعي المعياري تجد $z_{0.09}=2.23$.

بالتعويض في فترة الثقة المعطاة في نظرية(2) نجد :

$$73 - 2.33 \times \frac{6}{\sqrt{81}} < \mu < 73 + 2.33 + \frac{6}{\sqrt{81}}$$

وبالاختصار نحصل على :

$71.447 < \mu < 74.553$ فترة ثقة 98% للوسط μ.

مثال (10) :
إذا كان الوسط الحسابي والانحراف المعياري لمعدلات عينة عشوائية مؤلفة من 36 طالبا في احدى الجامعات هي $S = 12, \bar{X} = 87.3$.

أوجد (أ) فترة ثقة 95%.   (ب) فترة ثقة 99% لمعدل جميع طلبة الجامعة.

الحل: تباين المجتمع غير معلوم لكن n كبيرة، وبالتالي تستطيع استعمال النظرية (2) واستعمال S بدلا من σ.

(أ) $1 - \alpha = 0.95$ إذا $\frac{\alpha}{2} = 0.025$ إذاً $z_{0.975} = 1.96$

إذاً $$87.3 - 1.96 \times \frac{12}{\sqrt{36}} < \mu < 87.3 + 1.96 \times \frac{12}{\sqrt{36}}$$

وبالاختصار تكون فترة الثقة 95%

(83.38 , 91.22) أي $87.3 - 3.92 < \mu < 87.3 + 3.92$

(ب) $1 - \alpha = 0.99$ إذا $\frac{\alpha}{2} = 0.005$ إذا $z_{0.995} = 2.57$

إذاً $$87.3 - 2.57 \times \frac{12}{\sqrt{36}} < \mu < 87.3 + 2.57 \times \frac{12}{\sqrt{36}}$$

وبالاختصار $87.3 - 5.14 < \mu < 87.3 + 5.14$ أي (82.16 , 92.44)

وبمقارنة (أ)، (ب) فإننا نحتاج إلى فترة أطول إذا أردنا أن نعطي تقديرا أكثر ثقة للمعدل μ، ففي (أ) كان طول الفترة التي ثقتها 95% هو :

91.22 - 83.38 = 7.84

أما في (ب) فكان طول الفترة التي ثقتها 99 % هو :

92.44 - 82/16 = 10.28

8 : 8 فترة الثقة للوسط μ في حالة العينات الصغيرة

لقد وجدنا فترة الثقة للوسط μ عندما كانت المعاينة من مجتمع طبيعي تباينه σ^2 معلوم وكذلك عندما كانت المعاينة من أي مجتمع في حالة حجم العينة كبير.

أما عندما يكون حجم العينة صغيرا والتباين σ^2 غير معلوم ففي هذه الحالة فلا نستطيع تقريب $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ بالتوزيع الطبيعي المعياري، ولكن عندما تكون المعاينة من مجتمع طبيعي فإن توزيع $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ معلوم وقد درسناه سابقا. وهو توزيع t ذو درجات الحرية (n-1) ولذلك يمكننا استعمال :

النظرية (3) :

إذا أخذت عينة عشوائية حجمها n من مجتمع طبيعي فإن فترة ثقة $(1-\alpha)$100% للوسطμ تكون:

$$\left(\bar{X} - t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \bar{X} + t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right)$$

حيث $t_{\frac{\alpha}{2}}$ هي النقطة على محور t ذي درجات الحرية (n - 1) التي يعلوها $\frac{\alpha}{2}$ من المساحة ، أي أن

$$t_{\frac{\alpha}{2}} = t\left[1-\frac{\alpha}{2};n-1\right].$$

إن طريقة اشتقاق فترة الثقة في هذه النظرية هي نفسها طريقة اشتقاق فترة الثقة عندما تكون σ^2 معلومة ونوجزها كما يلي :

بما أن $\frac{\overline{X}-\mu}{S/\sqrt{n}}$ يخضع لتوزيع t ذي درجات الحرية (n -1) فإن :

$$P\left(\overline{X}-t_{\alpha/2}.\frac{S}{\sqrt{n}}<\mu<\overline{X}+t_{\alpha/2}.\frac{S}{\sqrt{n}}\right)=1-\alpha$$

وبالتالي نحصل على النظرية.

إن تفسير فترة الثقة السابقة هو نفس تفسير فترة الثقة عندما تكون σ^2 معلومة.

مثال (11) :

أخذت عينة عشوائية حجمها 9 من مجتمع طبيعي فأعطت $\overline{X} = 9.2$, S = 0.4، أوجد فترة ثقة 95% لمعدل المجتمع μ.

الحل : شروط النظرية متوفرة ولذلك نستعمل فترة الثقة.

$$\overline{X}-t_{\alpha/2}.\frac{S}{\sqrt{n}}<\mu<\overline{X}+t_{\alpha/2}.\frac{S}{\sqrt{n}}$$

بما أن $1 - \alpha = 0.95$ إذاً $\frac{\alpha}{2} = 0.025$ ومن جدول توزيع t بدرجات حرية 8 نجد أن $2.306 = t_{0.025}$.

إذن فترة الثقة 95 % للمعدل μ هي :

$$9.2-\frac{2.306\times 0.4}{\sqrt{9}}<\mu<9.2+\frac{2.306\times 0.4}{\sqrt{9}}$$

أي أن $9.20-0.307<\mu<9.20+0.307$

أي (9.51 , 8.89) هي فترة الثقة المطلوبة.

مثال (12) :

كانت محتويات 9 عبوات من أحد أنواع المنظفات كالآتي :

10.1 , 10.3 , 9.9 , 9.8 , 10.2 , 9.7 , 10.0 , 9.7 , 10.3 لترا.

أوجد فترة ثقة 99% لمعدل محتويات العبوات لذلك النوع من المنظفات على افتراض أن محتويات العبوات يخضع للتوزيع الطبيعي.

الحل : نحتاج لحساب معدل محتويات العبوات $\overline{X}$ والانحراف المعياري لتلك المحتويات S للعينة المعطاة.

$$\overline{X}=\frac{10.0+10.3+9.9+9.8+10.2+9.7+10.0+9.7+10.3}{9}$$

$$=10.0 \text{ لتراً}$$

$$S^2=\frac{\sum\left(X-\overline{X}\right)^2}{(n-1)}$$

$$S^2=\frac{(0.1)^2+(0.3)^2+(-0.1)^2+(-0.2)^2+(0.2)^2+(-0.3)^2+0+(0.3)^2}{8}$$

$$=\frac{0.01+0.09+0.01+0.04+0.04+0.09+O+0.09+0.09}{8}=0.057$$

إذن S = 0.24

بما أن $1-\alpha=0.99$ إذاً $\frac{\alpha}{2}=0.005$ إذن $t_{0.005}=3.355$ من جدول توزيع t ذي درجات الحرية 8

إذاً
$$10.0-3.355\times\frac{0.24}{\sqrt{9}}<\mu<10.0+3.355\times\frac{0.24}{\sqrt{9}}$$

وبالتالي فإن فترة الثقة هي : $9.73<\mu<10.27$

مثال (13) :

صنعت سبيكة لاستعمالها في أحد أنواع الصواريخ، وأخذت قياسات قوة السبيكة على 20 قطعة منها فوجد أن الوسط الحسابي 37.8 والانحراف المعياري 2.8.

(أ) أوجد فترة ثقة % 90 لمعدل قوة السبيكة.

(ب) هل تحوي هذه الفترة المعدل μ ؟

الحل : على فرض أن قوة السبيكة تخضع لتوزيع طبيعي وبتطبيق النظرية (3) فإن فترة الثقة هي :

$$\left(\overline{X} - t_{\alpha/2} \cdot \frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2} \cdot \frac{S}{\sqrt{n}}\right)$$

وبالتعويض حيث $t_{0.005} = 1.729$ لأن درجات الحرية 19

$$\left(37.8 - 1.729 \times \frac{2.8}{\sqrt{20}}, 37.8 + 1.729 \times \frac{2.8}{\sqrt{20}}\right)$$

وبالاختصار ، فالفترة هي : (38.88 , 36.72).

(ب) إننا لن نعلم فيما إذا كان فترة ثقة محددة مثل (38.88 , 36.72) تحوي μ أم لا لأن μ مجهولة ولكن لو كررنا العملية السابقة في (أ) لعدد كبير من العينات لأصبح عندنا ثقة أن % 90 من هذه الفترات يحوي μ وفقط %10 منها لا تحوي μ.

8 : 9 تقدير النسبة Estimation of Proportion

(1) تقدير النسبة بنقطة Point Estimation of Proportion

من المنطقي أن نقدر نسبة وجود ظاهرة في مجتمعٍ ما بنسبة وجود تلك الظاهرة في عينة عشوائية تؤخذ من ذلك المجتمع، وقد أيد ذلك تقدير النسبة بطريقة العزوم كما درسنا ذلك. فمثلا، إذا أردت أن تقدر نسبة العائلات التي تمتلك سيارة، فبإمكانك أن تختار عينة عشوائية ، وتحسب نسبة عدد العائلات التي تمتلك سيارة وتستعمل النسبة في العينة كتقدير نقطي للنسبة في المجتمع.

إن هذه المقدمة، تظهر لنا بوضوح أنه إذا كانت نسبة النجاح في تجربة ذات الحدين p يكون بالإمكان تقدير p كما يلي :

خذ عينة عشوائية حجمها **n** وافرض أن عدد النجاحات في هذه العينة **X**، يمكن استعمال $\overline{\mathbf{P}}=\frac{\mathbf{X}}{\mathbf{n}}$ كتقدير لنسبة النجاح p ويكون التقدير النقطي للمعلمة p هو نسبة النجاح في العينة $\overline{\mathbf{P}}=\frac{\mathbf{X}}{\mathbf{n}}$. لاحظ أن هذه العبارة تكافئ قولنا : خذ عينة عشوائية من مجتمع بيرنوللي $\mathbf{b(1,p)}$ ولتكن العينة ... $X_1, X_2,$ X_n , عندها يكون $X=\sum X_i$ هو عدد النجاحات في العينة ويكون $\overline{\mathbf{P}}=\frac{\mathbf{X}}{\mathbf{n}}=\frac{\sum \mathbf{X_i}}{\mathbf{n}}$.

لاحظ أن **X** هو الإحصاء : عدد النجاحات في العينة ، أما x فهو قيمة X ونحصل عليها من دراسة العينة أي أن x هو قيمة X التي نحصل عليها من عينة معينة.

والأمثلة على تجربة ذات الحدين كثيرة والأمثلة على الحاجة لتقدير نسبة النجاح p كثيرة أيضا، فمثلا ، تحتاج لتقدير نسبة الطلبة الذين يستعلمون النظارات الطبية في الصف العاشر الأساسي في بلد ما.

وتحتاج إلى تقدير نسبة الطلبة في الصف السادس الأساسي الذين يكتبون باليد اليسرى.

مثال (14) :
لتقدير نسبة المدخنين بين طلبة احدى الجامعات قام باحث بمقابلة عينة عشوائية حجمها 100 طالب فوجد أن 30 طالبا يدخنون، فما نسبة الطلبة المدخنين في الجامعة ؟

الحل : نسبة المدخنين في العينة $\mathbf{0.3}=\frac{\mathbf{30}}{\mathbf{100}}$. نقدر نسبة المدخنين في الجامعة بنسبة المدخنين في العينة وهي 0.3.

مثال (15) :

لمعرفة نسبة البيوت في مدينة ما التي يوجد فيها تدفئة مركزية ، أخذت عينة عشوائية حجمها 300 بيت ووجد أن 120 منها لديها تدفئة مركزية. ما هو التقدير النقطي لنسبة البيوت ذات التدفئة المركزية في تلك المدينة ؟

الحل : النسبة في العينة $\overline{P} = \frac{120}{300} = 0.40$

التقدير النقطي للنسبة p في المجتمع هي $\overline{P}$ أي 0.40

مثال (16) :

أردت معرفة نسبة المواطنين الذين يحبذون معالجة ماء الشرب بالكلور، فأخذت عينة حجمها 1000 مواطن فوجد أن 620 منهم يحبذون ذلك.

ما هو التقدير النقطي لنسبة المواطنين الحقيقية الذين يحبذون ذلك؟

الحل : النسبة في العينة $\overline{P} = \frac{620}{1000} = 0.62$ والتقدير النقطي للنسبة الحقيقية في المجتمع هو $\overline{P}$ أي 62 .0.

(2) تقدير النسبة بفترة Interval Estimation of Proportion

إن تقدير النسبة بفترة هو عبارة عن إيجاد تقدير نقطي لنسبة النجاح في المجتمع p ثم إيجاد توزيع المعاينة لذلك المقدر واستعمال هذه المعلومات لإيجاد فترة ذات معامل ثقة معين تحصر نسبة النجاح p داخلها.

إذا كان من المتوقع أن لا تكون نسبة النجاح غير المعلومة p قريبة جدا من الصفر أو الواحد، وكان حجم العينةn كبيرا فإن بإمكانك استعمال النظرية التي تفيد أن توزيع $Z = \frac{\overline{P} - p}{\sqrt{\frac{p(1-p)}{n}}}$ يقترب من التوزيع الطبيعي المعياري إذا كانت nكبيرة.

إذا توفرت هذه الشروط أمكنك وضع العبارة الاحتمالية

$$P\left(-Z_{1-\alpha/2} < Z < Z_{1-\alpha/2}\right) = 1 - \alpha$$

حيث $Z = \frac{\overline{P} - p}{\sqrt{\frac{p(1-p)}{n}}}$ تخضع للتوزيع الطبيعي المعياري تقريبا، $X, \overline{P} = \frac{X}{n}$ عدد النجاحات في العينة التي حجمها **n**.

من الصعب استعمال العبارة الاحتمالية السابقة بصيغتها المعطاة لإيجاد فترة ثقة للنسبة p، وذلك لأن p في المقام $\sqrt{\frac{p(1-p)}{n}}$ غير معلومة ، ولذلك نستعمل $\overline{P} = \frac{X}{n}$ بدلا من p في المقام فنحصل على فترة الثقة $(1-\alpha)$% 100 التقريبية للنسبة p وهي :

$$\overline{P} - z_{1-\alpha/2}\sqrt{\frac{\overline{P}(1-\overline{P})}{n}} < p < \overline{P} + z_{1-\alpha/2}\sqrt{\frac{\overline{P}(1-\overline{P})}{n}}$$

فترة الثقة للنسبة p.

نظرية (4) :

إذا كان $\overline{P} = \frac{X}{n}$ نسبة النجاح في عينة عشوائية حجمها n، وكان n كبيرا، فإن فترة الثقة $(1-\alpha)$% 100 التقريبية لنسبة النجاح p (p معلمة ذات الحدين ، نسبة النجاح في المجتمع) هي :

$$\overline{P} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{P}(1-\overline{P})}{n}} < p < \overline{P} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{P}(1-\overline{P})}{n}}$$

حيث $z_{1-\frac{\alpha}{2}}$ هي النقطة على محور الطبيعي المعياري الذي يقع إلى يسارها $\left(1-\frac{\alpha}{2}\right)$ من المساحة .

مثال (17) :

لإيجاد فترة ثقة % 95 لنسبة عدد الطلبة في المدارس الإعدادية الذين يستعملون النظارات الطبية، أخذت عينة عشوائية حجمها 400 طالب فوجد أن عدد مستعملي النظارات الطبية 100، أوجد فترة الثقة المطلوبة.

الحل :

نسبة النجاح في العينة تساوي $\overline{P} = \frac{100}{400} = \frac{1}{4}$

بما أن حجم العينة كبير يمكنك استعمال فترة الثقة التقريبية التالية، حيث $\alpha = 0.05$ لأن $1 - \alpha = 0.95$

$$\overline{P} - z_{0.975}\sqrt{\frac{\overline{P}(1-\overline{P})}{n}} < p < \overline{P} + z_{0.975}\sqrt{\frac{\overline{P}(1-\overline{P})}{n}}$$

بالتعويض تجد :

$$\frac{1}{4} - 1.96\sqrt{\frac{\frac{1}{4}\times\frac{3}{4}}{400}} < p < \frac{1}{4} + 1.96\sqrt{\frac{\frac{1}{4}\times\frac{3}{4}}{400}}$$
$$0.25 - 0.04 < p < 0.25 + 0.04$$

أي $0.21 < p < 0.29$

8 : 10 فترات الثقة للفرق بين وسطين والفرق بين نسبتين

Confidence Intervals for the difference between two means and the difference between two proportions

يمكننا استعمال الأسلوب السابق في بناء فترات الثقة للوسط وللنسبة لإيجاد فترات الثقة للفرق بين وسطين $(\mu_1 - \mu_2)$ والفرق بين نسبتين وذلك بالاستعانة بنظريات توزيعات المعاينة للفرق بين وسطين والفرق بين نسبتين.

نظرية (5) :

إذا كانت $X_1, X_2, \ldots, X_n$ عينة عشوائية من توزيع طبيعي $N(\mu_1, \sigma_1^2)$ وكانت $Y_1, Y_2, \ldots, Y_n$ عينة عشوائية من توزيع طبيعي $N(\mu_2, \sigma_2^2)$ مستقل عن التوزيع الأول ، وكانت σ_2^2, σ_1^2 معلومتين ، فإن فترة الثقة $100(1-\alpha)\%$ للفرق بين الوسطين $(\mu_1 - \mu_2)$ هي :

$$\left[(\bar{X}-\bar{Y}) - z_{1-\frac{\alpha}{2}}\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}\ ,\ (\bar{X}-\bar{Y}) + z_{1-\frac{\alpha}{2}}\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}\right]$$

مثال (18):

أخذت عينة عشوائية حجمها 16 من مجتمع طبيعي $N(\mu_1, 32)$ وأخذت عينة عشوائية حجمها 10 من مجتمع طبيعي $N(\mu_2, 70)$ مستقل عن الأول، فإذا أعطت العينة الأولى الوسط الحسابي $\bar{x} = 47$ وأعطت العينة الثانية الوسط الحسابي $\bar{y} = 35$ أوجد:

أ- فترة ثقة 95% للفرق بين الوسطين $(\mu_1 - \mu_2)$.

ب- فترة ثقة 90% للفرق بين الوسطين $(\mu_2 - \mu_1)$.

الحل: من الواضح أن شروط النظرية متحققة.

أ- $1-\alpha = 0.95$ إذاً $\alpha = 0.05$، $z_{0.975} = 1.96$ وتكون فترة الثقة

$$\left[(47-35) - 1.96\sqrt{\frac{32}{16}+\frac{70}{10}}, (47-35) + 1.96\sqrt{\frac{32}{16}+\frac{70}{10}}\right]$$

وبالاختصار تحصل على فترة الثقة :

$$[12 - 1.96 \times 3\ ,\ 12 + 1.96 \times 3]$$

أي $$[7.12, 17.88]$$

(ب) $1-\alpha = 0.90$ إذا $\alpha = 0.10\ ,\ z_{0.95} = 1.64$

وتكون فترة الثقة :

$$\left[(35-47)-1.64\sqrt{\frac{32}{16}+\frac{70}{10}}\ ,\ (35-47)+1.64\sqrt{\frac{32}{16}+\frac{70}{10}}\right]$$

وبالاختصار نحصل على فترة الثقة :

$$(-12-1.64\times 3\ ,\ -12+1.64\times 3)$$

أي $(-16.92\ ,\ -7.08)$

أما إذا لم يكن المجتمع الذي أخذنا منه العينة خاضعا للتوزيع الطبيعي فإن نظرية (5) تبقى متحققة شريطة أن يكون حجم كل عينة كبيرا أي أكثر أو يساوي 30 وذلك استنادا إلى النظرية الأساسية في تقارب التوزيعات.

نظرية (6) :

إذا كانت $X_1, X_2, \ldots, X_n$ عينة عشوائية من توزيع معدله μ_1 وتباينه σ_1^2 وكانت $Y_1, Y_2, \ldots, Y_n$ عينة عشوائية من توزيع مستقل عن الأول معدله μ_2 وتباينه σ_2^2 ، وكان σ_1^2, σ_2^2 غيرة معلومة فيمكن استعمال تباين كل عينة أي S_1^2 كتقدير لـ σ_1^2 ، S_2^2 كتقدير لـ σ_2^2، وتبقى النظرية متحققة شريطة أن n_2, n_1 كبيرتان.

مثال (19) :

دلت دراسات سابقة أن تباين علامات التوجيهي في احدى مدارس الذكور (أ) هو 169 وأن تباين علامات التوجيهي في احدى مدارس الإناث (ب) هو 144.

أخذت عينة عشوائية من مدرسة الذكور (أ) حجمها 85 طالباً فأعطت المعدل $\overline{X}=71$. وأخذت عينة عشوائية حجمها 72 من المدرسة (ب) فأعطت المعدل $\overline{Y}=67$ فإذا كان المعدل الحقيقي للعلامات في المدرسة (أ) هو μ_1 وفي المدرسة (ب) هو μ_2 أوجد فترة ثقة % 95 للفرق $\mu_1-\mu_2$.

الحل : المعطيات في المثال تحقق شروط النظرية ولذلك فإن فترة الثقة هي :

$$(71-67)\pm 1.96\sqrt{\frac{169}{85}+\frac{144}{72}}$$

أي $4\mp 1.96\times 2$ أي $(0.08\ ,7.92)$

ولكن عندما تكون تباينات المجتمعات غير معلومة وحجوم العينات صغيرة فيمكننا استعمال النظريات التالية :

نظرية (7) :

إذا كانت $X_1, X_2, \ldots, X_n$ عينة عشوائية من مجتمع طبيعي $N(\mu_1, \sigma_1^2)$ وكانت $Y_1, Y_2, \ldots, Y_n$ عينة عشوائية من مجتمع طبيعي $N(\mu_2, \sigma_2^2)$ مستقل عن الأول وكانت σ_1^2, σ_2^2 مجهولتين ولكن متساويتين فإن فترة ثقة $100(1-\alpha)\%$ للفرق بين الوسطين $(\mu_1 - \mu_2)$ هي :

$$(\overline{X} - \overline{Y}) \pm t_{\frac{\alpha}{2}} . S_c . \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

حيث S_c^2 هو التباين المجمَّع للعينتين معا، أي :

$$S_c^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

وأن $t_{\alpha/2}$ هي النقطة على محور t ذي درجات الحرية $(n_1 + n_2 - 2)$ والتي تحصر إلى يمينها $\frac{\alpha}{2}$ من المساحة (أي $t_{\alpha/2} = t\left[1 - \frac{\alpha}{2}, n_1 + n_2 - 2\right]$).

مثال (20) :

أخذت عينة عشوائية حجمها 9 من الأطفال حديثي الولادة في المستشفى (أ) في احدى المدن فأعطت معدل الأوزان $\overline{X} = 2.7$ ، $S_1^2 = 1.2$ وأخذت عينة عشوائية حجمها 15 من الأطفال حديثي الولادة في مستشفى (ب) في الريف فأعطت معدل الأوزان $\overline{Y} = 3.1$ ، $S_2^2 = 1.4$، فإذا كان معدل أوزان الأطفال المولودين في (أ) هو μ_1 والمعدل للأطفال المولودين في (ب) هو μ_2 فأوجد فترة ثقة 90% للفرق بين الوسطين $(\mu_2 - \mu_1)$.

الحل : نفرض أن أوزان الأطفال حديثي الولادة في كل من المستشفيين (أ)، (ب) يخضع للتوزيع الطبيعي ونفرض أن تباين المجتمع الأول σ_1^2 يساوي تباين المجتمع الثاني σ_2^2 (مع أنهما مجهولان) ومن وجهة عامة فيمكن اختبار هذا الغرض كما سندرسهُ في بند قادم ولكن نكتفي هنا بفرض التساوي إذا كان $2 \le S_1^2 / S_2^2 \le \frac{1}{2}$.

بعد هذا الفرض فإن شروط النظرية (1) متحققة ولذلك نجد أولا S_c^2 حيث:

$$S_c^2 = \frac{(9-1)1.2 + (15-1)14}{9+15-2} = 1.38$$

$$S_c = \sqrt{1.38} = 1.17$$

$1-\alpha = 0.90$ إذا $\frac{\alpha}{2} = 0.05$ ويكون

$$t_{\frac{\alpha}{2}} = t[0.95,22] = 1.717$$

وتكون فترة الثقة %90 هي :

$$(3.1-2.7) \mp 1.717 \times 1.17 \times \sqrt{\frac{1}{9} + \frac{1}{15}}$$

أي

$0.4 \mp 1.717 \times 1.17 \times 0.42$ أي 0.4 ∓ 0.84

أي أن الفترة المطلوبة (‎-0.44 , 1.24‎).

وكما حصلنا على فترة ثقة نسبة النجاح في المجتمع فإننا نحصل على فترة ثقة للفرق بين نسبتين شريطة ن يكون حجوم العينات كبيرة وذلك كما في النظرية :

نظرية (8) :

إذا كانت $X_1, X_2, \dots, X_n$ عينة عشوائية من مجتمع بيرنوللي $b(1, p_1)$ وكانت $Y_1, Y_2, \dots, Y_n$ عينة عشوائية من مجتمع مستقل عن الأول $b(1, p_2)$ فإن فترة ثقة $100(1-\alpha)\%$ للفرق بين النسبتين $(p_1 - p_2)$ هي :

$$(\overline{P}_1 - \overline{P}_2) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\overline{P}_1(1-\overline{P}_1)}{n_1} + \frac{\overline{P}_1(1-\overline{P}_2)}{n_2}}$$

شريطة أن n_2, n_1 كبيرتان ، وحيث $\overline{P}_1, \overline{P}_2$ هما نسبتا النجاح في العينتين على التوالي.

مثال (21) :
أخذت عينة عشوائية حجمها 100 من طلبة الصف السادس الأساسي في احدى القرى (أ) فوجد أن 27 منهم يستعملون النظارات الطبية، وفرضنا أن النسبة الحقيقية p_1 وأخذت عينة عشوائية حجمها 80 من طلبة الصف السادس الأساسي في المدينة (ب) فوجد أن 32 منهم يستعملون النظارات الطبية وفرضنا أن النسبة الحقيقية p_2.

أوجد فترة ثقة % 95 للفرق $(p_2 - p_1)$.

الحل : حجما العينتين كبيران.

باستعمال النظرية (8)، $\overline{P}_1 = 0.27$ ، $\overline{P}_2 = \frac{32}{80} = 0.4$

فترة الثقة هي :

$$(0.4 - 0.27) \pm 1.96\sqrt{\frac{0.27 \times 0.73}{100} + \frac{0.4 \times 0.6}{80}}$$

وبالاختصار نجد أن الفترة هي :

$$0.13 \pm 1.96 \times 9 \times 0.07$$

أي (0.269 , 0.008-)

8 : 11 حجم العينة Sample Size

بعد دراستنا لكيفية إيجاد فترات الثقة، ينشأ لدينا السؤال التالي : قبل إجراء التجربة وتسجيل قيم المشاهدات في العينة، ما هو حجم العينة المطلوبة كي نحصل على فترة ثقة بطول معلوم أو بخطأ محدد لا نريد أن نتجاوزه.

لاحظ أنه إذا كانت المعاينة من توزيع طبيعي تباينه معلوم أو من توزيع ما تباينه معلوم وحجم العينة كبير فإن فترة الثقة $100(1-\alpha)\%$ للوسط μ هي :

$$\left[\overline{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$$

$\overline{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ $\quad \overline{X} \quad \bar{x} \quad \mu \quad$ $\overline{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

وطول هذه الفترة يساوي $2\,z_{1-\frac{\alpha}{2}}\,\frac{\sigma}{\sqrt{n}}$ فإذا أردنا طول الفترة لا يزيد عن d فذلك يعني :

$$2z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le d$$

إذن

$$n \ge \left(\frac{2z_{1-\frac{\alpha}{2}}\sigma}{d}\right)^2$$

أي أن حجم العينة المطلوب هو أصغر عدد صحيح n يحقق المتباينة أعلاه.

مثال (22) :

أوجد حجم العينة n التي يجب أخذها من توزيع طبيعي $N(\mu,18)$ للحصول على فترة ثقة 95% للوسط وبطول لا يزيد عن 1.6.

الحل : نأخذ أصغر عدد n بحيث

$$n \ge \left(\frac{2 \times z_{1-\frac{\alpha}{2}}\sigma}{d}\right)^2$$

أي

$$n \ge \left(\frac{2 \times 1.96 \times \sqrt{18}}{1.6}\right)^2 = 108.04$$

إذا نأخذ n = 109

لاحظ أن الخطأ في التقدير هو $\left|\overline{X}-\mu\right|$ وبالنظر إلى فترة الثقة ورسمها البياني، فلو كانت μ في داخل الفترة على يمين μ أو على يسارها فمن الواضح أن الخطأ = $\left|\overline{X}-\mu\right| \leq z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$. إن

المقدار $z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ يسمى الحد الأعلى للخطأ

ولذلك إذا كان الخطأ المسموح به e يكون لدينا

$$e \leq z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

إذن

$$n \geq \left(z_{1-\frac{\alpha}{2}} \frac{\sigma}{e} \right)^2$$

مثال (33) :

في المثال السابق ما قيمة n بحيث أن الخطأ لا يتجاوز 0.5.

الحل :

$$n \geq \left(\frac{1.96 \times \sqrt{18}}{0.5} \right)^2 = 276.57$$

إذن

n = 277

وبنفس الطريقة يمكننا إيجاد حجم العينة إذا كان المجتمع مجتمع بيرنوللي b (1, p) ، وفي هذه الحالة ومن نظرية (4) يكون طول فترة الثقة $2Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p(1-p)}{n}}$ ،

وبما أن p غير معلومة ولا نستطيع تقديرها بالقيمة $\overline{P}$ لأننا لم نأخذ العينة بعد، ولذلك نضع بدلا من p (1-p) أكبر قيمة لها وهي $\frac{1}{4}$ لأن الاقتران p (1-p) على الفترة (0, 1) له قيمة عظمى عند $p=\frac{1}{2}$ والقيمة العظمى تكون $\frac{1}{4}$.

إذا أكبر قيمة لطول الفترة هي :

$$2z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{1/4}{n}} = z_{1-\frac{\alpha}{2}} \cdot / \sqrt{n}$$

فإذا أردنا أن هذا الطول لا يتجاوز d نحصل على $z_{1-\frac{\alpha}{2}} \cdot / \sqrt{n} \leq d$

إذن $n \geq \left(\frac{z_{1-\frac{\alpha}{2}}}{d} \right)^2$

وإذا أردنا أن الخطأ في التقدير لا يتجاوز e (لاحظ d = 2e = طول الفترة).

نحصل على $n \geq \left(\frac{z_{1-\frac{\alpha}{2}}}{2e} \right)^2$

مثال (24) :

ما حجم العينة من مجتمع b (1, p) التي يجب أخذها للحصول على فترة ثقة % 95 بطول لا يتجاوز 0.2.

الحل : $n \geq \left(\frac{z_{1-\frac{\alpha}{2}}}{d} \right)^2 = \left(\frac{1.96}{0.2} \right)^2$

$$= 96.04$$

إذا $n = 97$

مثال (25) :

ما حجم العينة من مجتمع b (1 , p) للحصول على فترة ثقة % 90 بخطأ لا يزيد عن 0.05.

الحل :

$$n \geq \left(\frac{z_{1-\frac{\alpha}{2}}}{2e} \right)^2 = \left(\frac{1.64}{2 \times 0.05} \right)^2$$

$$= 268.96$$

إذا $n = 269$

8 : 12 فترات الثقة للتباين Confidence Intervals for Variance

لاحظنا في الفصل السابع أنه إذا كانت $X_1, X_2, \ldots, X_n$ عينة عشوائية من مجتمع طبيعي $N(\mu, \sigma^2)$ فإن $(n-1)S^2/\sigma^2$ تخضع لتوزيع χ^2 بدرجات حرية (n-1)، حيث S^2 هي تباين العينة,

لهذا لإيجاد فترة ثقة $100(1-\alpha)\%$ للتباين σ^2 فإننا نعلم من جداول توزيع χ^2 أن النقطتين $\chi^2\left[1-\frac{\alpha}{2}; n-1\right], \chi^2\left[\frac{\alpha}{2}; n-1\right]$ تحصران بينهما مساحة $(1-\alpha)$ تحت توزيع χ^2 بدرجات حرية (n-1) أي أن :

$$P\left(\chi^2\left[\frac{\alpha}{2}; n-1\right] \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2\left[1-\frac{\alpha}{2}; n-1\right] \right) = 1-\alpha$$

وبالتبسيط الجبري (خذ مقلوب كل حد مع تغيير اتجاه ا لمتباينة ثم اضرب بـ $(n-1)S^2$ نجد :

$$P\left(\frac{(n-1)S^2}{\chi^2\left[1-\frac{\alpha}{2}; n-1\right]} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2\left[\frac{\alpha}{2}; n-1\right]} \right) = 1-\alpha$$

وهذا يؤدي إلى النظرية :

نظرية (9) :

إذا كانت $X_1, X_2, \ldots, X_n$ عينة عشوائية من توزيع طبيعي $N(\mu, \sigma^2)$ فإن فترة $100(1-\alpha)\%$ ثقة للتباين σ^2 هي :

$$P\left(\frac{(n-1)S^2}{\chi^2\left[1-\frac{\alpha}{2};n-1\right]} \ , \ \frac{(n-1)S^2}{\chi^2\left[\frac{\alpha}{2};n-1\right]} \right)$$

حيث S^2 هي تباين العينة $S^2 = \sum\left(X_i - \overline{X}\right)^2 /(n-1)$

$$= \frac{\sum X_i^2 - n\overline{X}^2}{(n-1)}$$

ونجد فترة الثقة للانحراف المعياري بإيجاد الجذر التربيعي لجذور فترة الثقة للتباين.

مثال (26) :

عينة عشوائية حجمها 20 أخذت من مجتمع طبيعي $N\left(\mu,\sigma^2\right)$ فأعطت التباين $S^2 = 15$ أوجد فترة ثقة 95% للتباين σ^2 .

الحل: $1-\alpha = 0.95$ إذا $\frac{\alpha}{2} = 0.025$ ، $1-\frac{\alpha}{2} = 0.975$ ، درجات الحرية 19= 1 - 20

$$\chi^2[0.025;19] = 8.907$$
$$\chi^2[0.975;19] = 32.852$$

وحسب النظرية فإن فترة الثقة هي :

$$\left[\frac{19\times 15}{32.852} \ , \ \frac{19\times 15}{8.907}\right]$$

أي هي : $[8.675 \, , 31.997]$

أما فترة ثقة % 95 للانحراف المعياري فهي : $\left[\sqrt{8.675} \, , \sqrt{31.997}\right]$ أي هي $[2.94 \, , 5.66]$.

8 : 13 فترات الثقة للنسبة بين تباينين

Confidence intervals for the ratio of two variances

حسب النظرية [(9) الفصل السابع] فإن $\frac{S_1^2}{S_2^2}\frac{\sigma_2^2}{\sigma_1^2}$ يخضع للتوزيع F_{n_1-1,n_2-1}. لهذا إذا أردنا إيجاد فترة $100(1-\alpha)\%$ ثقة للنسبة $\frac{\sigma_2^2}{\sigma_1^2}$ أي أردنا إيجاد الإحصائيين L , U بحيث أن

$$P\left(L \le \frac{\sigma_2^2}{\sigma_1^2} \le U\right) = 1-\alpha$$

فإننا نعود إلى توزيع F ونلاحظ أن :

$$P\left(F\left[\frac{\alpha}{2};n_1-1,n_2-1\right] \le F \le F\left[1-\frac{\alpha}{2};n_1-1,n_2-1\right]\right) = 1-\alpha$$

وبالتعويض $\frac{S_1^2}{S_2^2}\frac{\sigma_2^2}{\sigma_1^2}$ بدلا من F نجد

$$P\left(F\left[\frac{\alpha}{2};n_1-1,n_2-1\right] \le \frac{S_1^2}{S_2^2}\frac{\sigma_2^2}{\sigma_1^2} \le F\left[1-\frac{\alpha}{2};n_1-1,n_2-1\right]\right) = 1-\alpha$$

وبالضرب بـ $\frac{S_2^2}{S_1^2}$ نجد :

$$P\left(\frac{S_2^2}{S_1^2}F\left[\frac{\alpha}{2};n_1-1,n_2-1\right] \le \frac{\sigma_2^2}{\sigma_1^2} \le \frac{S_2^2}{S_1^2}F\left[1-\frac{\alpha}{2};n_1-1,n_2-1\right]\right) = 1-\alpha$$

وبالتالي نحصل على النظرية:

نظرية (10) :

إذا كانت $X_1, X_2, \dots, X_n$ عينة عشوائية من $N(\mu_1, \sigma_1^2)$ وكانت

$Y_1, Y_2, \ldots, Y_n$ عينة عشوائية من $N(\mu_2, \sigma_2^2)$ مستقل عن المجتمع الأول فإن فترة $100(1-\alpha)\%$ ثقة ، للنسبة σ_2^2/σ_1^2 هي :

$$\left(\frac{S_2^2}{S_1^2}F\left[\frac{\alpha}{2};n_1-1,n_2-1\right] \ , \ \frac{S_2^2}{S_1^2}F\left[1-\frac{\alpha}{2};n_1-1,n_2-1\right]\right)$$

مثال (27) :

أخذت عينة عشوائية حجمها $n_1 = 9$ من مجتمع $N(\mu_1, \sigma_1^2)$ فأعطت التباين $S_1^2 = 65.4$ وأخذت عينة عشوائية حجمها $n_2 = 12$ من مجتمع $N(\mu_2, \sigma_2^2)$ مستقل عن الأول فأعطت التباين $S_2^2 = 127.3$. أوجد فترة 90% ثقة للنسبة σ_2^2/σ_1^2 .

الحل : شروط النظرية متحققة ، $1-\alpha = 0.90$ إذاً $\frac{\alpha}{2} = 0.05$ ، ومن جداول توزيع F نجد :

$$F[0.05;8,11] = \frac{1}{F[0.95;11,8]} = \frac{1}{3.31} = 0.30$$
$$F[0.95;8.11] = 2.95$$

إذاً فترة الثقة بالتعويض في نظرية (5) هي:

$$\left[\frac{127.3}{65.4}\times 0.3 \ , \ \frac{127.3}{65.4}\times 2.95\right]$$
$$[0.588 \ , \ 5.742]$$

تمارين

1-8 : أخذت عينة عشوائية حجمها 12 من مجتمع وسطه μ وكانت أفراد العينة : 12, 3, .10, 9, 8, 6, 8, 4, 7, 5, 3, 6

فما هو تقدير μ ؟

2-8 : أردت تقدير نسبة الطلبة الذين يستعملون النظارة الطبية في إحدى المدارس الأساسية في العاصمة، فأخذت عينة عشوائية حجمها 20 وسجلت "نعم" لكل طالب يستعمل النظارة و "لا" لكل طالب لا يستعملها فكانت النتيجة :

لا , لا , لا , نعم , لا , لا , نعم , لا , نعم , نعم ، لا , لا , لا , لا , لا , نعم , لا ، لا ، نعم ، لا.

ما تقديرك لنسبة الذين يستعملون النظارة ؟

3-8 : لتقدير نسبة عدد الطلبة في إحدى الجامعات الذين رسبوا في ثلاثة مقررات أو أكثر خلال دراستهم ، قام أحد الباحثين بدراسة عينة عشوائية حجمها 900 طالب فوجد أن 243 منهم قد رسبوا في ثلاثة مقررات أو أكثر.

أ) قدر نسبة عدد الطلبة في الجامعة الذين رسبوا في ثلاثة مقررات أو أكثر.

ب) أوجد فترة ثقة 95% لهذه النسبة.

4-8 : أخذت عينة عشوائية حجمها 400 من معلمي المرحلة الاعدادية فوجد أن 80 منهم حاصلون على شهادة البكالوريوس.

أ) قدر نسبة المعلمين في المرحلة الاعدادية الحاصلين على شهادة البكالوريوس.

ب) أوجد فترة ثقة 99% للنسبة الحقيقية للمعلمين في هذه المرحلة الحاصلين على شهادة البكالوريوس.

5-8 : ينتج أحد المصانع مصابيح كهربائية تخضع أعمارها تقريبا لتوزيع طبيعي انحرافه المعياري 35 ساعة.

أخذت عينة عشوائية حجمها 25 مصباحا، فكان الوسط الحسابي لأعمار هذه المصابيح 890 ساعة، أوجد فترة ثقة 98% لمعدل أعمار المصابيح التي ينتجها المصنع؟

6-8 : أخذت عينة من 60 خيطا من مصنع للخيوط فوجد أن معدل قوة هذه الخيوط 90.4كغم بانحراف معياري مقداره 7.5كغم، أوجد فترة ثقة 95% لمعدل قوة جميع الخيوط التي ينتجها ذلك المصنع.

7-8 : أخذت عدة أفلام من إنتاج شركة ما بطريقة عشوائية فوجد أن الفترة الزمنية لها هي :

101 , 98 , 103 , 105 , 96 , 99 , 102 97 ، 104 دقيقة.

أ) أوجد تقديرا نقطيا لمعدل زمن جميع الأفلام التي أنتجتها تلك الشركة.

ب) على فرض أن أزمان الأفلام تخضع لتوزيع طبيعي ، أ وجد فترة ثقة 95% لمعدل زمن جميع الأفلام التي أنتجتها تلك الشركة ؟

حـ) هل هناك مبرر لافتراض التوزيع الطبيعي؟ استعمل طريقة رسم الاحتمالات الطبيعية من الفصل السادس.

8-8 : لتقدير معدل المصروف الشهري بالدينار لدى طلاب احدى الجامعات قام باحث بمقابلة عينة عشوائية من الطلبة حجمها 25 فوجد أن مصروفهم الشهري كما يلي :

38.0, 51.7, 49.2, 38.3, 36.5, 37.8, 48.1, 50.2, 35.5, 37.0, 42.0, 43.6,
41.0, 52.3, 49.4, 50.3, 30.2, 38.7, 32.6, 41.8, 33.9, 35.1, 44.0, 39.0.

ما هو تقديرك لمعدل المصروف الشهري لجميع طلبة الجامعة ؟

أوجد فترة ثقة 90% لمعدل المصروف إذا فرضنا أن المصروف يخضع لتوزيع طبيعي.

9-8 : إذا كانت نسبة العائلات التي لديها هاتف في مدينة ما p أخذت عينة عشوائية حجمها 50 عائلة في هذه المدينة، فكانت نسبة من لديهم هاتف في هذه العينة 0.65 ، أوجد فترة ثقة 95% للنسبة p.

10-8 : أراد باحث تقدير معدل كمية الفوسفات لكل وحدة حجم من ماء بحيرة. إذا علم أن الانحراف المعياري حسب دراسات سابقة تقريبا ثابت وهو ملغم4= σ ما هو عدد المشاهدات (يحصل الباحث على مشاهدة من كل عينة) التي يجب أن يجريها الباحث ليكون واثقاً 95% أن الخطأ في التقدير لا يزيد عن 0.7 ملغم ؟.

8-11 : لإيجاد فترة ثقة 95% لمعدل مجتمع، أ وجد حجم العينة المطلوبة بحيث أن (أ) لا يزيد الخطأ عن $\frac{1}{9}$ الانحراف المعياري للمجتمع. (ب) لا يزيد الخطأ عن 15% من الانحراف المعياري للمجتمع.

8-12 : لتقدير μ بفترة ثقة 90% وخطأ 2.5 وحدة، وجد باحث أن حجم العينة يجب أن يكون 108. ما هو حجم العينة المطلوب إذا أراد الباحث فترة ثقة 95% بخطأ لا يزيد عن 1.8 وحدة؟

8-13 : رش أحد الباحثين 150 ذبابة بمبيد حشري خفيف وسجل فترة حياتها فوجد أن الوسط 12.5 دقيقة والانحراف المعياري 3.4 دقيقة.

أوجد فترة ثقة 99% لمعدل فترة الحياة الحقيقية.

8-14 : سجل مزارع عدد الأيام التي استغرقتها بذور البازلاء للإنبات فكانت 20 ,19 ,12 ,8 ,17 ,13 ,17 ,11 ,18 ,.

أوجد فترة ثقة 95% لمعدل عدد الأيام التي يستغرقها هذا النوع من البازلاء للإنبات. أذكر الفرضيات التي تحتاجها لتتمكن من الحل.

8-15 : سجلت قياسات الحموضة (PH) لعينات من ماء المطر في 10 مواقع في منطقة صناعية فكانت ،3.6 ,4.1 , 3.9 , 4.8 , 3.2 , 4.5 , 3.8 , 5.1 , 3.1 , 3.9 .

أوجد فترة ثقة 90% لمعدل حموضة ماء المطر في تلك المنطقة.

8-16 : في تمرين (8-14) أوجد فترة ثقة 90% لتباين عدد الأيام للإنبات.

8-17 : في تمرين (8-15) أوجد فترة ثقة 95% لتباين معدل حموضة ماء المطر.

8-18 : أخذت عينة عشوائية حجمها 5 من مجتمع طبيعي $N(\mu,\sigma^2)$ فأعطت $S^2 = 17$.
أوجد فترة ثقة 95% للتباين σ^2. أوجد فترة ثقة 95% للانحراف المعياري σ.

8-19 : لمعرفة نسبة الطلبة الذين يكتبون باليد اليسرى في المدارس الأساسية أخذت عينة عشوائية من تلك المدارس حجمها 1000 فوجد أن 80 منهم يكتبون باليد اليسرى.

أ) قدر النسبة الحقيقية للطلبة الذين يكتبون باليد اليسرى.

ب) أوجد فترة ثقة %90 للنسبة الحقيقية للطلبة الذين يكتبون باليد اليسرى.

20-8 : لتقدير نسبة المصابين باللشمانيا في احدى المدن، أوجد حجم العينة المطلوب للحصول على فترة ثقة %95 لتلك النسبة بحد أقصى للخطأ مقداره 0.1.

21-8 : في استطلاع شمل 625 طالباً من الناجحين في شهادة الدراسة الثانوية العامة وجد أن %30 منهم قد التحقوا بالجامعة، %25 قد التحقوا بكليات المجتمع، ولم يكمل الباقون دراستهم بعد الثانوية العامة.

(أ) أوجد فترة ثقة %95 للنسبة الحقيقية للطلبة الملتحقين بالجامعة ؟

(ب) أوجد فترة ثقة %90 للنسبة الحقيقية للطلبة الذين لم يواصلوا دراستهم بعد الثانوية العامة.

22-8 : للمقارنة بين أعداد البكتيريا في وحدة الحجم في مياه بحيرتين أخذت باحثة عينة عشوائية حجمها 10 وحدات من البحيرة (أ) فوجدت أعداد البكتيريا :

175 , 207 , 200 , 195 , 205 , 193 , 225 , 205 , 97 , 198.

وأخذت عينة أخرى من البحيرة (ب) فوجدت الأعداد :

205 , 217 , 210 , 125 , 203 , 212 , 208.

على فرض أن عدد البكتيريا في كل من البحيرتين يخضع لتوزيع طبيعي وأن التباين في المجتمعين σ^2، أوجد :

(أ) فترة ثقة %95 لمعدل عدد البكتيريا الحقيقي في البحيرة (أ).

(ب) فترة ثقة %90 لمعدل عدد البكتيريا الحقيقي في البحيرة (ب).

(جـ) فترة ثقة %95 للفرق بين المعدلين في البحيرتين (أ) ، (ب).

23-8 : أخذت عينة عشوائية حجمها 9 من توزيع طبيعي $N(\mu_1, \sigma_1^2)$ فأعطت $\overline{X} = 13.5$ ، $S_1^2 = 5$ بينما أعطت عينة عشوائية حجمها 13 مأخوذة من مجتمع $N(\mu_2, \sigma_2^2)$ مستقل عن الأول ، وسطا حسابيا $\overline{Y} = 12.2$ ، $S_2^2 = 4.5$.

أ) أوجد فترة ثقة %95 للتباين σ_1^2.

ب) أوجد فترة ثقة 90% للانحراف المعياري σ_2.

جـ) أوجد فترة ثقة 95% للنسبة σ_2^2 / σ_1^2.

8-24 : أخذت عينة عشوائية حجمها 60 من علامات الطالبات في المدرسة (أ) فأعطت $\overline{X} = 69$ ، $S_1^2 = 144$ وأخذت عينة عشوائية حجمها 80 من علامات الطلاب الذكور في المدرسة (ب) فأعطت $\overline{Y} = 66$ ، $S_2^2 = 169$.

أوجد فترة ثقة للفرق بين المعدلين الحقيقيين لعلامات الطالبات في (أ) وعلامات الطلاب في (ب).

8-25 : إذا كانت أجور مندوبي المبيعات الذكور تخضع لتوزيع طبيعي $N(\mu_1, 100)$ ، وأجور مندوبات المبيعات تخضع لتوزيع طبيعي $N(\mu_2, 144)$ مستقل عن الأول.

أخذت عينة عشوائية حجمها 10 مندوبين فأعطت الوسط الحسابي $\overline{X} = 170$, $S_1 = 9$ وأخذت عينة عشوائية حجمها 24 من المندوبات فأعطت الوسط الحسابي $\overline{Y} = 162$ ، $S_2 = 10$.

(أ) أوجد فترة ثقة 95% لكل من μ_1, μ_2.

(ب) أوجد فترة ثقة 95% للفرق $(\mu_1 - \mu_2)$.

8-26 : للمقارنة بين تحصيل الطلبة في القرية والمدينة عقد باحث امتحانا تحصيليا لعينات عشوائية من مدارس القرية ومدارس المدينة فكانت النتائج كما يلي :

المدارس	حجم العينة	الوسط الحسابي	الانحراف المعياري
المدينة	200	79	11
القرية	150	73	12

أوجد فترة ثقة 95% للفرق بين معدلي تحصيل الطلبة في المدينة والقرية.

الفصل التاسع
اختبار الفرضيات

Testing Hypotheses

إن أحد فروع الإحصاء الاستنتاجي هو اختبار الفرضيات، فنحن في كثير من الأحيان لا نكتفي بتقدير معلمة المجتمع بأن نعطيها قيمة معينة أو نبني لها فترة ثقة معينة، بل نحتاج إلى اتخاذ قرار حول صحة فرضية معينة أو عدم صحتها، أي أننا نحتاج إلى اختبار الفرضيات المتعلقة بمعلمات المجتمع. وهذا ما سنقوم بدراسته بشكل مبسط حيث سندرس اختبار الفرضيات البسيطة والتي تمت تسميتها هكذا لأنها تتعلق باختبار فرضية محدّدة عن معلمة واحدة مثل الوسط الحسابي أو نسبة النجاح.

9 : 1 الفرضية الصفرية والفرضية البديلة

Null Hypothesis and Alternative hypothesis

تعريف (1) :
الفرضية الاحصائية هي كل عبارة عن إحدى معلمات المجتمع تكون قابلة للاختبار وبالتالي تكون صحتها أو عدم صحتها بحاجة إلى قرار.

من هنا فإن الفرضية الاحصائية تتعلق بعبارة عن احدى المعلمات مثل

وسط المجتمع أو نسبة النجاح أو التباين أو غيرها.

مثال (1) :
إذا كان معدل طول الجندي في أحد الجيوش حوالي 167سم، فالجمل التالية تعتبر فرضيات حول هذا المعدل :

العبارة كلاما	العبارة رياضيا
معدل الطول 167 سم	$H : \mu = 167$
معدل الطول أقل من 167 سم	$H : \mu < 167$
معدل الطول أكثر من 167 سم	$H : \mu > 167$
معدل الطول لا يساوي 167 سم	$H : \mu \neq 167$

في معظم الأحيان هناك نوعان من الفرضيات في المسألة الواحدة. النوع الأول هو الفرضية الصفرية وهي الفرضية التي تبنى على أمل أن يتخذ قرار بعدم صحتها، ونصطلح الآن على اعتبار أي فرضية نود اختبارها فرضية صفرية، ونعبر عن ذلك بالرمز H_0. وهكذا ، فكل فرضية احصائية تريد اختبارها تسمى فرضية صفريةH_0 .

إن رفض الفرضية الصفرية H_0 يؤدي إلى قبول فرضية أخرى تسمى الفرضية البديلة، وتعبر عنها بالرمز H_1.

وعندما تصاغ الفرضية الصفرية المتعلقة بمعلمة مجتمع معين بشكل يعيّن قيمة محددة لتلك المعلمة فإنها تسمى فرضية بسيطة. ففي المثال (1) تعتبر الفرضية : معدل الطول 167سم أي 167 = $H_0 : \mu$ فرضية صفرية بسيطة، وهي كما تلاحظ تعين قيمة محددة للمعلمة μ هي 167سم.

أما الفرضية البديلة فيمكن أن تعين قيمة محددة للمعلمة تحت الدراسة كما تسمح بأن تأخذ المعلمة قيما متعددة. ارجع إلى مثال (1) تجد أن الفرضية :

معدل الطول أكبر من 167سم أي 167 $H_1 : \mu$ < تعتبر فرضية بديلة، وهي كما تلاحظ تعطي إمكانية أن يأخذ المعدل أي قيمة أكبر من 167، فقد يكون 171 أو 169 أو أي قيمة أخرى أكبر من 167سم.

9 : 2 عملية اختبار الفرضيات الإحصائية

Process of testing statistical Hypotheses

القيم الحرجة Critical Values

كيف نختبر الفرضيات الإحصائية ؟ أي ما هي القواعد لعملية اتخاذ القرار برفض أو عدم رفض الفرضية الصفرية H_0 لصالح الفرضية البديلة H_1.

سنبدأ بشرح اختبار الفرضية الصفرية $H_0 : \mu = 167$.

مقابل الفرضية البديلة $H_1 : \mu \neq 167$ حيث μ هي معدل طول الجندي في أحد الجيوش كما في المثال (1).

إن الفرضية في هذا المثال تتعلق بمعدل المجتمع ولذلك كان من المنطقي أن نبني اختبار هذه الفرضية باستعمال معدل العينة (الوسط الحسابي لها).

من نظرية تقارب التوزيعات نعلم أنه إذا كان $\mu = 167$ هو معدل المجتمع σ^2 تباينه فإن توزيع المعاينة للوسط الحسابي $\overline{X}$ للعينات الكبيرة الحجم سيكون قريبا من التوزيع الطبيعي الذي معدله 167 وتباينه $\frac{\sigma^2}{n}$ حيث n حجم العينة.

وهذا يعني أنه إذا استعملت عينة عشوائية كبيرة من المجتمع تحت الدراسة وتفحصت هذه العينة لوجدت أن هناك فرصة صغيرة جدا لوقوع وسطها الحسابي $\overline{X}$ على بعد أكثر من انحرافين معياريين $2\sigma_{\bar{x}}$ أو ثلاثة انحرافات معيارية $3\sigma_{\bar{x}}$ عن المعدل المفروض $\mu = 167$ (لاحظ أن احتمال وقوع $\overline{X}$ ضمن انحرافين معياريين من المعدل $\mu = 167$ هو 0.9544 وبالتالي فإن احتمال وقوع $\overline{X}$ على بعد أكثر من انحرافين معياريين لا يصل إلى 0.05.

والآن ، إذا حدث وحصلت على قيمة متطرفة للوسط الحسابي $\overline{X}$ بحيث كانت هذه القيمة على بعد أكثر من انحرافين معياريين من $\mu = 167$ فهناك حالتان : إما أن حادثا نادر الحدوث قد وقع أي أنك قد اخترت عينة من تلك العينات نادرة الحدوث وغير الممثلة للمجتمع، أو أن معدل المجتمع لا يساوي القيمة التي حددتها في فرضيتك الصفرية $\mu = 167$. ستختار الحالة الثانية وتستعمل تلك النتيجة (كون القيمة $\overline{X}$ بعيدة أكثر من انحرافين معياريين عن $\mu = 167$) كقاعدة لرفض الفرضية الصفرية.

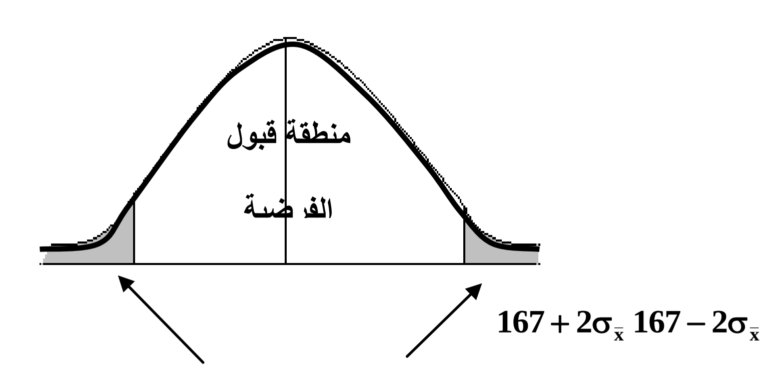
في المثال السابق ، تم استعمال $\overline{X}$ لنبني على قيمته رفض أو عدم رفض الفرضية حول وسط المجتمع μ، أما إذا كانت المعلمة المطلوب اختبارها (نسبة النجاح p) أي أن الفرضية الصفرية كانت متعلقة بنسبة النجاح p فإننا نستعمل نسبة النجاح في العينة $\overline{P}$ لنبني عليها الاختبار. ومن وجهة عامة فإننا نستعمل إحصاء مناسبا لنبني عليه اختبار الفرضية تحت الدراسة، ويسمى هذا الاحصاء (إحصاء الاختبار). ففي الفرضية حول μ فإن إحصاء الاختبار هو $\overline{X}$. وإذا كانت الفرضية عن p (نسبة النجاح في المجتمع) فإن إحصاء الاختبار هو $\overline{P}$ (نسبة النجاح في العينة) وهكذا.

تعريف (2) :

إحصاء الاختبار Test Statistic

هو إحصاء (اقتران تعين قيمته من العينة) يبني عليه قرار اختبار الفرضيات. وهكذا فإن $\overline{X}$ هو إحصاء اختبار وكذلك $\overline{P}$ ، S^2.

وبالعودة إلى مثالنا في اختبار الفرضية الصفرية $H_0 : \mu = 167$ مقابل الفرضية البديلة $H_1 : \mu \neq 167$، ورد معنا أن من المنطقي أن نرفض H_0 إذا كانت القيمة $\overline{X}$ (الوسط الحسابي للعينة) على بعد أكثر من انحرافين معياريين $2\sigma_{\overline{x}}$ من الوسط $\mu = 167$ ، وذلك لأن احتمال وقوع $\overline{X}$ على بعد أكثر من $2\sigma_{\overline{x}}$ من الوسط $\mu = 167$ صغير جدا (أقل من 0.05). وتسمى المنطقة إلى يمين (أعلى من) $2\sigma_{\overline{x}}$ + 167 أو على يسار (أصغر من) $2\sigma_{\overline{x}}$ -167 منطقة الرفض. أما الفترة $(167 - 2\sigma_{\overline{x}}, 167 + 2\sigma_{\overline{x}})$ فتسمى منطقة قبول الفرضية $H_0 : \mu = 167$ ، أنظر الشكل (1).



شكل (1)

وكما هو ملاحظ فإنه يتم تعيين منطقة الرفض ومنطقة القبول إذا ما عرفت $\sigma_{\bar{x}}$. ففي مثالنا السابق، إذا كان الانحراف المعياري لأطوال الجنود هو $\sigma = 6$ سم وإذا أخذت عينة حجمها n =36 فإن

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ وبالتالي $\sigma_{\bar{x}} = \frac{6}{\sqrt{36}} = 1$.

وتكون منطقة الرفض : إلى يمين 169 = 1 × 2 + 167 أو إلى يسار 165 = 1 × 2 - 167
وتكون منقطة القبول (169، 165).

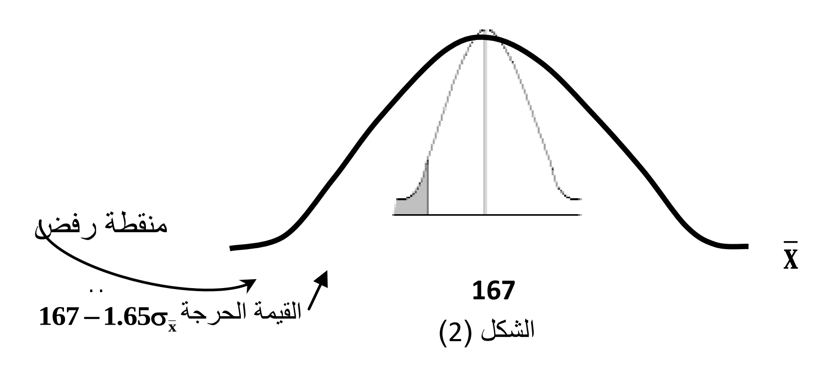
فإذا كان وسط العينة التي تم اختيارها $\overline{X} = 170$ مثلا فإنك ترفض الفرضية $H_0 : \mu = 167$.

أما إذا كان $\overline{X} = 168$ مثلا فإنك لا ترفض الفرضية H_0.

ويسمى العددان 165 ، 169 القيمتين الحرجتين. أما منطقة رفض H_0 فتسمى المنطقة الحرجة.

تعريف (3) :
المنطقة الحرجة للاختبار هي مجموعة قيم إحصاء الاختبار التي تؤدي إلى رفض الفرضية الصفرية ، كل حد من حدود المنطقة الحرجة يسمى قيمة حرجة لإحصاء الاختبار.

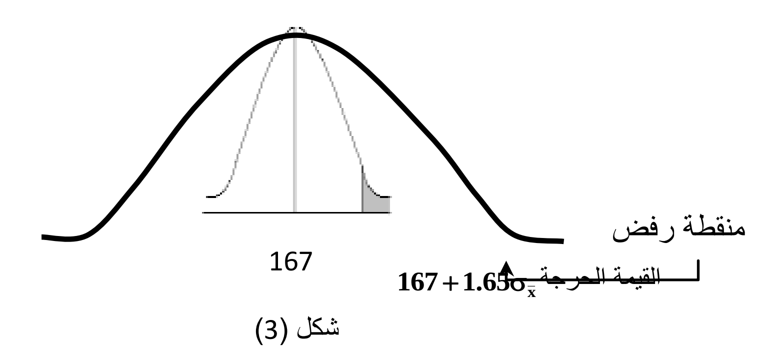
لاحظ أنه تم تعيين قيمتين حرجتين لأن الفرضية البديلة كانت $H_1 : \mu \neq 167$، أما لو كانت الفرضية البديلة $H_1 : \mu < 167$ فإن القيمة الحرجة تكون النقطة اليسرى فقط وتكون المنطقة الحرجة هي المنطقة الواقعة إلى يسار القيمة الحرجة واستعملنا 1.65 بدلاً من 2 لنحصل على المساحة المظللة في الجهة اليسرى أقل من 0.05.



الشكل (2)

أما إذا كانت $H_1: \mu > 167$ فإن القيمة الحرجة تكون النقطة اليمنى.

وتكون المنطقة الحرجة : جميع قيم $\overline{X}$ إلى يمين تلك النقطة واستعملنا 1.65 لتكون المساحة المظللة أقل من 0.05.



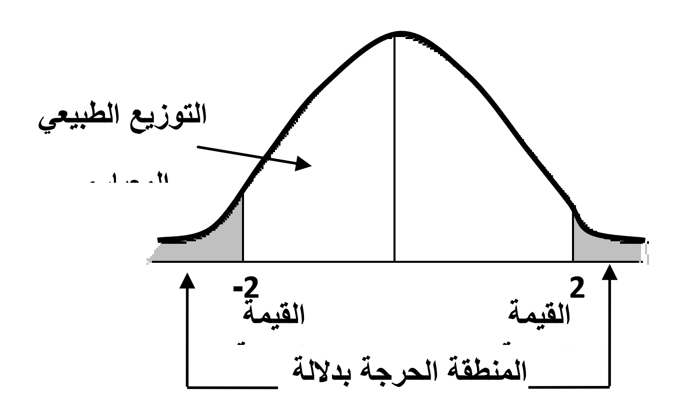
شكل (3)

لقد عيّنت المنطقة الحرجة والقيم الحرجة لاختبار الفرضية الصفرية $H_0: \mu = 167$ بدلالة إحصاء الاختبار $\overline{X}$ ، ومن المعلوم أن توزيع $\overline{X}$ يقرب من الطبيعي إذا كان حجم العينة n كبيرا أو كان المجتمع خاضعا لتوزيع الطبيعي تباينه معلوم لذلك يمكنك تعيين المنطقة الحرجة والقيم الحرجة بدلالة إحصاء الاختبار Z الذي تحصل عليه من $Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$ شريطة أن تكون σ معلومة.

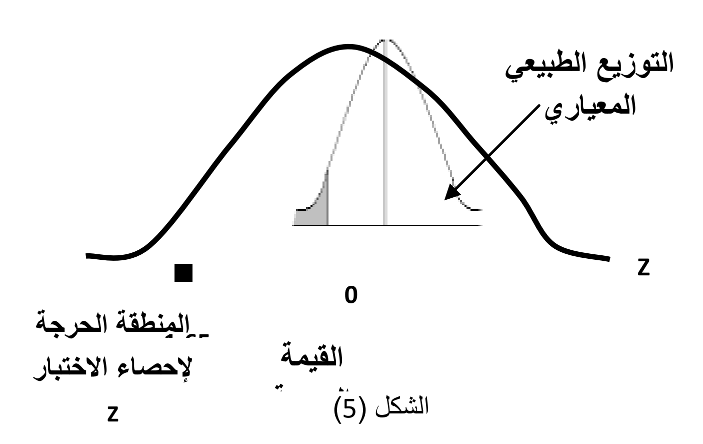
لذلك فإن القيم الحرجة بدلالة Z المقابلة للقيم الحرجة بدلالة $\overline{X}$ في المثال السابق هي :

$$Z = \frac{165 - 167}{6 / \sqrt{36}} = -2$$

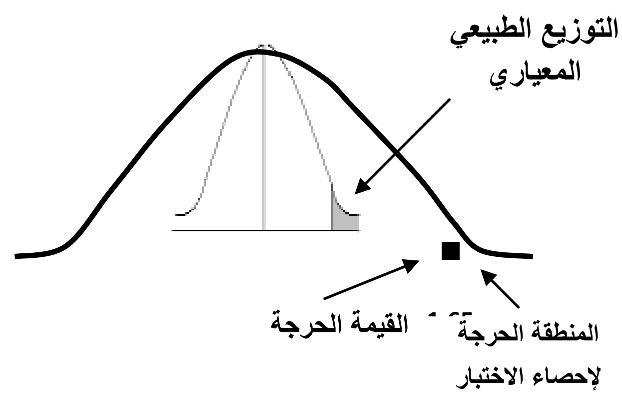
$$Z = \frac{169 - 167}{6 / \sqrt{36}} = 2$$



الشكل (4)



الشكل (5)



الشكل (6)

إن المنطقة الحرجة في كل من الأشكال (2)، (3)، (5)، (6) تقع في جهة واحدة وإن الاختبار في كل منها : اختبار ذو طرف واحد أما القيم الحرجة في الشكلين (1)، (4) فهي قيم حرجة باتجاهين، وهي تعني أن المنطقة الحرجة باتجاهين ويكون الاختبار في هذه الحالة : اختبارا ذا طرفين.

9 : 3 الخطأ من النوع الأول والخطأ من النوع الثاني

Type I error and Type II error

كل قرار يبنى على نتائج عينة يكون معرضا للخطأ، حيث أن مثل هذه القرارات تبنى على متغير عشوائي أو إحصاء مثل $\bar{P}, \bar{X}$ وغيرها. وفي اختبار الفرضيات، هناك حالتان بالنسبة للفرضية الصفرية، هما : إما أن تكون هذه الفرضية صحيحة وإما أن تكون غير صحيحة. وهناك أيضا نوعان من القرارات التي يتخذها الإحصائي بصدد الفرضية الصفرية، هما : رفض الفرضية الصفرية أو عدم رفضها.

فإذا كانت H_0 صحيحة وكان قرارنا رفضها، فمعنى ذلك الوقوع في خطأ. وهو يسمى في هذه الحالة : الخطأ من النوع الأول.

أما إذا كانت الفرضية الصفرية غير صحيحة وكان القرار عدم رفضها، فمعنى ذلك الوقوع في خطأ أيضا، لكنه في هذه الحالة يسمى : الخطأ من النوع الثاني.

تعريف (4) :
يحدث الخطأ من النوع الأول إذا رفضت الفرضية الصفرية وهي في الحقيقة صحيحة.

تعريف (5) :

يحدث الخطأ من النوع الثاني إذا لم ترفض الفرضية الصفرية ، وهي في الحقيقية غير صحيحة (أنظر الجدول).

قرار الإحصائي	الة المعلمة	H_0 صحيحة	H_0 غير صحيحة (H_1 صحيحة)
	اقبل H_0	قرار صحيح	خطأ من النوع الثاني
	ارفض H_0	**طأ من النوع الأول**	قرار صحيح

من تعاريف الأخطاء من النوع الأول والنوع الثاني نستطيع حساب احتمالات حدوثها كما يلي :

تعريف (6) :

احتمال الخطأ من النوع الأول، ونعبر عنه بالرمز α هو :

$\alpha = P$ (H_0 صحيحة | رفض الفرضية H_0).

أي : α تساوي احتمال رفض الفرضية H_0 إذا علم أن H_0صحيحة.

احتمال الخطأ من النوع الثاني، ونعبر عنه بالرمز β هو :

$\beta = P$ (H_1 صحيحة | عدم رفض الفرضية H_0).

أي : β تساوي احتمال عدم رفض الفرضية H_0 إذا علم أن H_1صحيحة.

عند حساب الاحتمالات السابقة لاحظ أن الحادث "رفض الفرضية H_0 " يعني وقوع إحصاء الاختبار في منطقة الرفض، وأن العبارة " H_0 صحيحة" تعني استعمال قيمة المعلمة المحددة لك في الفرضية الصفرية.

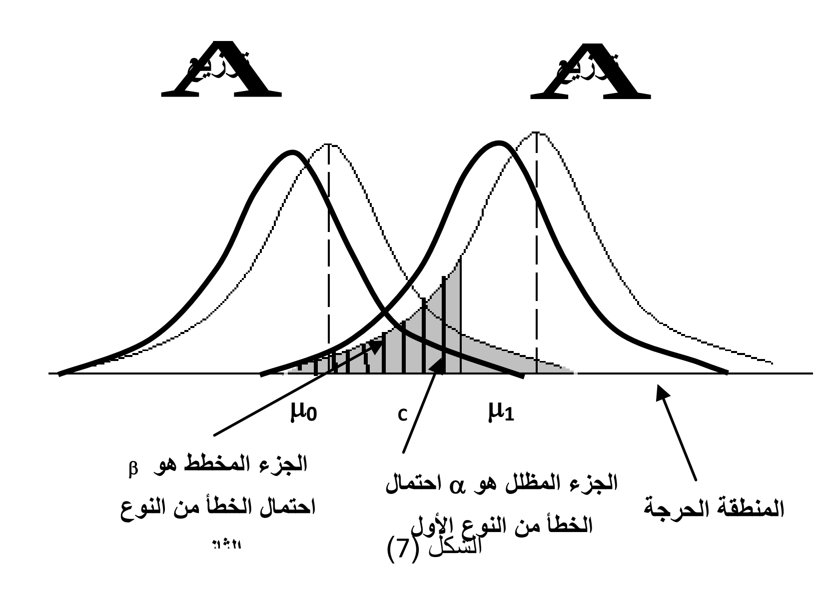
إن الحادث "عدم رفض H_0" يعني وقوع إحصاء الاختبار في متممة منطقة الرفض، أي متممة المنطقة الحرجة.

والعبارة " H_1صحيحة" تعني وجوب إعطاء قيمة معينة للمعلمة تحت الاختبار على أن تكون هذه القيمة من القيم المسموح بها تحت الفرضية البديلة H_1.

ويمكن توضيح الخطأ من النوع الأول والخطأ من النوع الثاني والاحتمالات β, α بالرسم كما يظهر في الشكل (7) الذي يمثل اختبارا حول وسط المجتمع μ، حيث الفرضيات هي :

$H_0 = \mu = \mu_0$

$H_1 = \mu > \mu_0$ وأخذنا قيمة محدّدة هي μ_1



الشكل (7)

9 : 4 اختبار الفرضيات المتعلقة بالوسط

Testing Hypotheses Concerning Mean

كان اختبار الفرضية في المثال (1) مبنيا على قرار رفض الفرضية الصفرية بناء على مفهوم منطقي يدعو إلى رفض تلك الفرضية $H_0 : \mu = \mu_0$ إذا كان الوسط الحسابي $\overline{X}$ بعيدا عن μ_0 بأكثر من انحرافين معياريين للوسط.

والسؤال الآن :

كيف نجري الاختبار إذا لم تستعمل الاختبار السابق المبني على المفهوم المنطقي؟

قبل الإجابة عن هذا السؤال نحاول الإجابة عن السؤال : هل يستطاع ضمان احتمالات صغيرة β,α للخطأ من النوع الأول والخطأ من النوع الثاني ؟

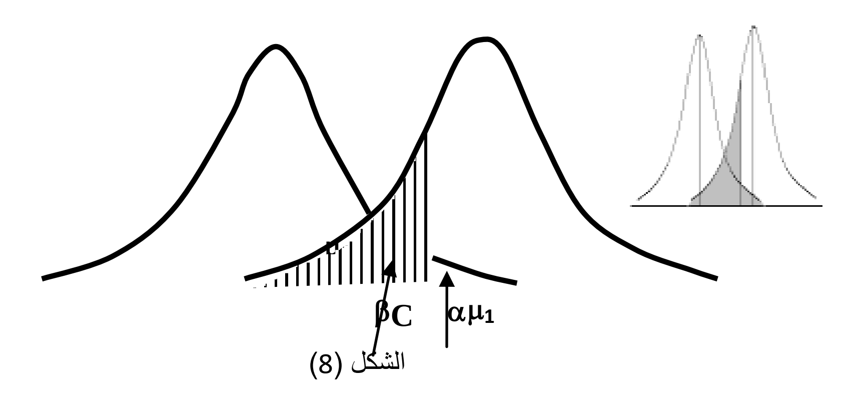
ولذلك سندرس العلاقة بين β,α في اختبار الفرضية الصفرية :

$$H_0 : \mu = \mu_0$$

مقابل الفرضية البديلة $H_1 : \mu > \mu_1$.

افرض أن القيمة الحرجة C وأن المنطقة الحرجة : ارفض H_0 إذا كان $\overline{X} > C$.

أنظر الشكل (8).



الشكل (8)

من تعريف الخطأ من النوع الأول والخطأ من النوع الثاني وبالنظر إلى الشكل يتضح أن احتمال الخطأ من النوع الأول α هو مساحة الجزء المظلل، أما احتمال الخطأ من النوع الثاني المتعلق بالقيمة μ_1 , حيث $\mu_1 > \mu_0$ فهو β ويساوي مساحة الجزء المخطط في الشكل ويتضح كذلك أنه إذا تحركت C إلى اليمين صغرت قيمة α وكبرت قيمة β، وبالعكس إذا تحركت C إلى اليسار ازدادت قيمة α وصغرت قيمة β.

إذن لا يمكن جعل قيمتي α,β صغيرتين في نفس الوقت بواسطة تغيير المنطقة الحرجة. يتضح أيضا أننا نستطيع حساب قيمة α في معظم اختبارات الفرضيات لأنها تحسب باعتبار قيمة μ قيمة معلومة هي μ_0 . أما β فعلينا أن نحسبها لكل قيمة من قيم μ حسب تغير قيمة μ في منطقة الفرضية H_1.

مما سبق، يبدو منطقيا أن تستعمل الاختبارات التي تتمكن من جعل احتمال الخطأ من النوع الأول α صغيرا ولكنه قيمة ثابتة ونحاول جعل β صغيرة. ومن هذا المنطلق ، تبقى α صغيرة، ونجد المنطقة الحرجة التي تحقق ذلك ، وبالتالي نبني اختباراً للفرضية الصفرية ونكون متأكدين أن احتمال الخطأ من النوع الأول فيه صغير وقيمته α. في مثل هذه الأحوال ندعو α مستوى الدلالة للاختبار.

تعريف (7) :

مستوى الدلالة لاختبار ما ، هو احتمال الخطأ من النوع الأول α، ويعبر عنه كنسبة مئوية ونحدّده قبل إجراء الاختبار.

وهكذا ، إذا كان $\alpha = 0.05$ نقول مستوى الدلالة للاختبار 5%.

وعادة أن يكون مستوى الدلالة 1% أو 5% وهما الأكثر استعمالا في اختبار الفرضيات ويكون بناء الاختبار ذي مستوى الدلالة α:

أولا : بتحديد وجود طرف واحد للاختبار أو طرفين ، أي وجود قيمة حرجة واحدة أو قيمتين ويعتمد هذا على الفرضية البديلة. فإذا كانت الفرضية البديلة من النوع $H_1 : \mu < \mu_0$ أو $H_1 : \mu > \mu_0$ كان للاختبار قيمة حرجة واحدة، أي كان الاختبار ذا طرف واحد ؛ وإذا كانت الفرضية البديلة من النوع $H_1 : \mu \neq \mu_0$ كان هناك قيمتان حرجتان وكان للاختبار طرفان.

ثانيا : بعد ذلك، نجد القيمة الحرجة (القيم الحرجة) بحيث يكون احتمال الخطأ من النوع الأول يساوي α، أي أن المساحة في أقصى يمين أو يسار القيمة الحرجة تساوي α في حالة الاختبار ذي الطرف الواحد. ويكون مجموع المساحتين في أقصى يمين القيمة الحرجة اليمنى وفي أقصى يسار القيمة الحرجة اليسرى يساوي α في حالة الاختبار ذي الطرفين لكل طرف $\frac{\alpha}{2}$ علما بأن هذه المساحات تحسب على أساس معرفة التوزيع الاحتمالي لإحصاء الاختبار.

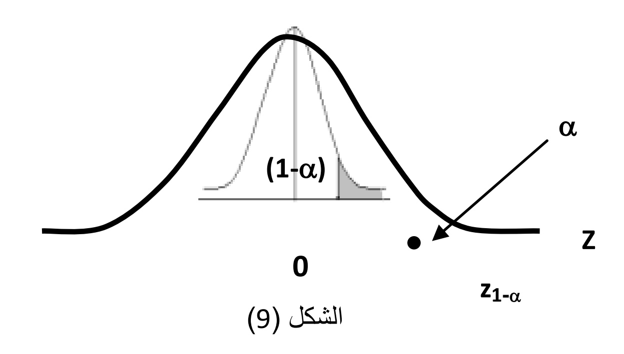
عند استعمال Z كإحصاء اختبار فرضية معينة ، على مستوى الدلالة α فإن القيم الحرجة تكون :

$z_{1-\alpha}$ لاختبار الفرضية $H_0 : \mu = \mu_0$ إذا كانت الفرضية البديلة $H_1 : \mu > \mu_0$.

z_{α} لاختبار الفرضية H_0 إذا كانت الفرضية البديلة $H_1 : \mu < \mu_0$.

$+z_{1-\alpha/2}$ ، $-z_{1-\alpha/2}$ لاختبار الفرضية $H_0 : \mu = \mu_0$ إذا كانت الفرضية البديلة $H_1 : \mu \neq \mu_0$.

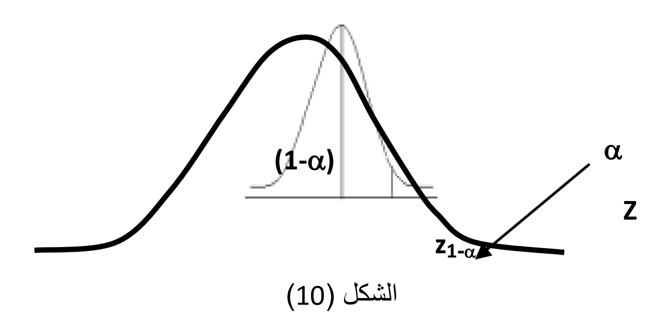
حيث $z_{1-\alpha}$ هي القيمة على المحور الأفقي للتوزيع الطبيعي المعياري الذي يكون إلى يسارها مساحة قيمتها $(1-\alpha)$ أي يكون إلى يمينها مساحة α.



الشكل (9)

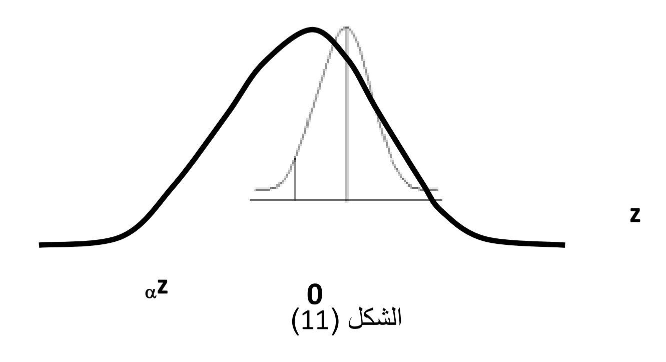
أما القيم الحرجة بدلالة $\overline{X}$ فيمكن حسابها بعد إيجاد القيم الحرجة بدلالة Z وأنت تحتاج في هذه الحال إلى معرفة الإنحراف المعياري للمجتمع σ والذي يمكن الاستعاضة عنه باستعمال الانحراف المعياري للعينة S إذا كانت σ غير معلومة وحجم العينة كبيرا.

الأشكال التالية توضح القيم الحرجة بدلالة Z :



الشكل (10)

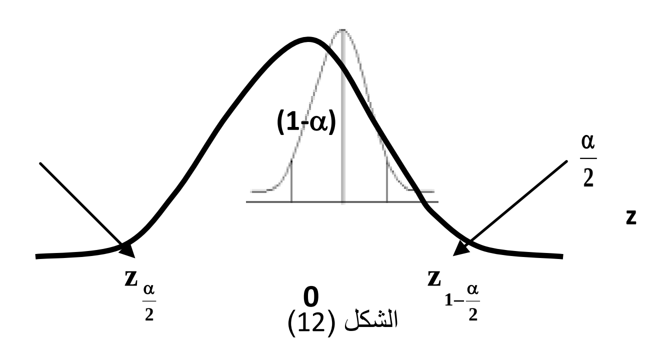
1- القيمة الحرجة $z_{1-\alpha}$ للاختبار ذي الطرف الواحد للفرضية البديلة $\mu > \mu_0$ بمستوى الدلالة α.



الشكل (11)

2- القيمة الحرجة $+\mathbf{z}_{\alpha}$ للاختبار ذي الطرف الواحد للفرضية البديلة $\mu < \mu_0$ بمستوى الدلالة α
أي القيمة الحرجة هي $z_{1-\alpha}$-بسبب التماثل.

3- القيم الحرجة $\mathbf{z}_{\frac{\alpha}{2}}, \mathbf{z}_{1-\frac{\alpha}{2}}$ للاختبار ذي الطرفين للفرضية البديلة $\mu \neq \mu_0$ بمستوى دلالة α لاحظ أن $\mathbf{z}_{\alpha/2} = -\mathbf{z}_{1-\alpha/2}$ بسبب التماثل.



الشكل (12)

9 : 5 اختبـار الفرضـيات المتعلقـة بالوسـط لمجتمـع طبيعـي تباينه معلوم

Testing Hypotheses Concerning the mean of Normal Distribution with Known Variance

لقد بحثنا معظم الأفكار الأساسية حول اختبار الفرضيات المتعلقة بالوسط الحسابي تحت فرض هام جدا، وهو : استعمال التوزيع الطبيعي كتقريب لتوزيع وسط العينة $\overline{\mathbf{X}}$.

والآن سنبحث النظريات التي تستعمل كقاعدة لتبرير استعمال إحصاء الاختبار المناسب في كل حالة.

نظرية (1) :

إذا أخذت عينة عشوائية حجمها n من مجتمع طبيعي $N(\mu,\sigma^2)$ وكانت σ^2 معلومة فإن إحصاء الاختبار للفرضية $H_0 : \mu = \mu_0$ هو $Z = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$ وهو يخضع للتوزيع الطبيعي المعياري، حيث $\overline{X}$ هو الوسط الحسابي للعينة.

خطوات الاختبار :

(1) اختبر $H_0 : \mu = \mu_0$.

(2) مقابل الفرضية البديلة :

(i) $H_1 : \mu > \mu_0$

أو (ii) $H_1 : \mu < \mu_0$

أو (iii) $H_1 : \mu \neq \mu_0$

(3) مستوى الدلالة α.

(4) إحصاء الاختبار : تحت فرض ان H_0 صحيحة فإن إحصاء الاختبار هو $Z = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$ يخضع لتوزيع طبيعي معياري.

(5) القيم الحرجة ومنطقة الرفض:

ارفض الفرضية H_0 إذا كان

(i) $Z > z_{1-\alpha}$ بالنسبة للفرضية البديلة $H_1: \mu>\mu_0$.

(ii) $Z < z_{\alpha}$ بالنسبة للفرضية البديلة $H_1: \mu<\mu_0$.

(iii) $Z > Z_{1-\alpha/2}$ أو $Z < Z_{\alpha/2}$ بالنسبة للفرضية البديلة $H_1: \mu\neq\mu_0$.

(6) الحسابات: احسب Z المعرّف في (4).

(7) قارن Z التي حسبتها في (6) مع منطقة الرفض في (5) حسب الفرضية المعطاة لديك وقرر رفض H_0 أو عدم رفضها.

ويمكن إجراء الاختبار باستعمال القيم الحرجة بدلالة $\bar{x}$ وذلك باستعمال الخطوات (1) - (4) السابقة كما هي ونكمل عليها الخطوات التالية:

(5) القيم الحرجة : بالنسبة للفرضية البديلة (i) ، القيمة الحرجة $z_{1-\alpha}$ ويقابلها $\bar{X}_1$ من المعادلة :

$$z_{1-\alpha} = \frac{\bar{x}_1 - \mu_0}{\sigma/\sqrt{n}}$$

بالنسبة للفرضية البديلة (ii) ، القيمة الحرجة z_{α} ويقابلها $\bar{x}_2$ من المعادلة:

$$z_{\alpha} = \frac{\bar{x}_2 - \mu_0}{\sigma/\sqrt{n}}$$

بالنسبة الفرضية البديلة (iii)، فإن الاختبار المناسب ذو الطرفين، والقيم الحرجة هي :

$$z_{\alpha/2} , z_{1-\alpha/2}$$

ويقابلها : $\bar{x}_2$, $\bar{x}_1$ من المعادلتين :

$$z_{\alpha/2} = \frac{\bar{x}_2 - \mu_0}{\sigma/\sqrt{n}}$$

$$z_{1-\alpha/2} = \frac{\bar{x}_1 - \mu_0}{\sigma/\sqrt{n}}$$

(6) الحسابات :

ولإجراء الاختبار احسب $\bar{X}$ من العينة التي اخترتها.

(7) المقارنة : قارن هذه القيمة بالقيمة الحرجة حسب الفرضية البديلة المعطاة.

الحالة (i) : $H_1 : \mu > \mu_0$ ارفض H_0 إذا كان $\bar{X} > \bar{x}_1$.

الحالة (ii) : $\mathbf{H_1 : \mu < \mu_0}$ ارفض H_0 إذا كان $\overline{\mathbf{X}} < \overline{\mathbf{x}}_2$.

الحالة (iii) : $\mathbf{H_1 : \mu \neq \mu_0}$ ارفض H_0 إذا كان $\overline{\mathbf{X}} > \overline{\mathbf{x}}_1$ أو $\overline{\mathbf{X}} < \overline{\mathbf{x}}_2$.

مثال (8) :

تخضع أوزان عبوات أحد مساحيق الغسيل لتوزيع طبيعي انحرافه المعياري 7غم ومعدلة μ.

على مستوى دلالة 0.05 اختبر الفرضية :

$\mathbf{H_0 : \mu = 200}$ مقابل الفرضية البديلة،

$\mathbf{H_1 : \mu \neq 200}$ ، إذا كان الوسط الحسابي لعينة حجمها 25 علبة هو $\overline{\mathbf{X}} = \mathbf{208}$ غم.

الحل :

(1) $\mathbf{H_0 : \mu = 200}$

(2) $\mathbf{H_1 : \mu \neq 200}$

(3) $\boldsymbol{\alpha = 0.05}$

(4) إحصاء الاختبار هو : تحت فرض أن H_0 صحيحة فإن : $\mathbf{Z = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}}$ يخضع لتوزيع طبيعي معياري.

(5) المنطقة الحرجة : بما أن الفرضية البديلة لا تحدد اتجاها واحدا، فإن الاختبار المناسب ذو طرفين، والقيم الحرجة تكون :

$$\mathbf{Z_{0.025} = -1.96 \;,\; Z_{0.975} = 1.96}$$

والمنطقة الحرجة هي : ارفض H_0 إذا كان $Z > 1.96$.

أو $Z < -1.96$ أنظر الشكل (13).

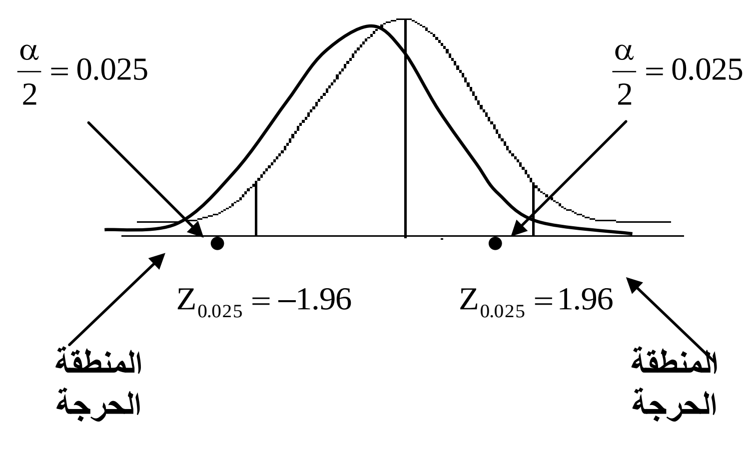
(6) الحسابات : احسب Z من المعادلة $\mathbf{Z = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}}$

أي :

$$\mathbf{Z = \frac{208 - 200}{7/\sqrt{25}} = \frac{8 \times 5}{7} = 5.5}$$

(7) المقارنة : قارن قيم Z بالقيمة الحرجة :

من الواضح أن 1.96 < 5.5 أي أن قيمة Z تقع في المنطقة الحرجة، لذلك ارفض H_0 لصالح 200 <μ ولاحظ أن النتيجة كانت لصالح 200 <μ لأن Z وقعت في منطقة الرفض اليمنى، أي على يمين $\mathbf{Z_{1-\alpha/2}}$.



الشكل (13)

مثال (9) :
يخضع الزمن الذي يحتاجه الطالب للتسجيل في احدى الجامعات إلى توزيع طبيعي انحرافه المعياري 0.7 ساعة ومعدله μ ساعة.

اختبر الفرضية $\mathbf{H_0 : \mu = 5.2}$ ساعة.

مقابل الفرضية البديلة $\mathbf{H_1 : \mu > 5.2}$

أ- باستعمال القيمة الحرجة بدلالة $\mathbf{\overline{X}}$.

ب- باستعمال القيمة الحرجة بدلالة z.

على مستوى دلالة 0.05 إذا أعطت عينة حجمها 16 وسطا حسابيا $\overline{X} = 5.3$ ساعة.

الحل :

أ -

(1) $H_0 : \mu = 5.2$

(2) $H_1 : \mu > 5.2$

(3) $\alpha = 0.05$

(4) بما أن الفرضية البديلة ذات اتجاه واحد هو "أكبر من" فإن الاختبار المناسب ذو طرف واحد والقيمة الحرجة $Z_{0.05} = 1.645$ ، والمنطقة الحرجة: ارفض H_0 إذا كان $Z > 1.645$.

(5) إن القيمة الحرجة بدلالة $\overline{X}$ هي :

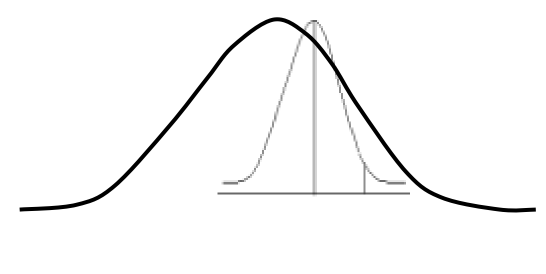
$$1.645 = \frac{\bar{x} - 5.2}{0.7/\sqrt{16}}$$

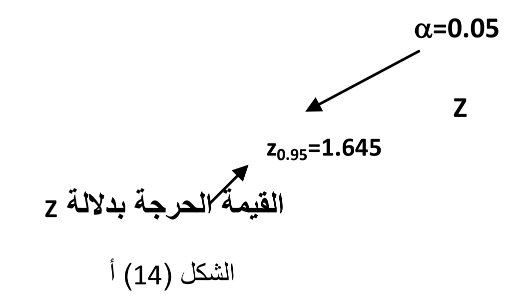
إذن $$\bar{x} = 5.2 + \frac{1.645 \times 0.7}{4}$$

= 5.49

والمنطقة الحرجة: ارفض H_0 إذا كان $\overline{X} > 5.49$.

(6) قارن قيمة $\overline{X}$ التي حصلت عليها من العينة مع القيمة الحرجة $\bar{x}_1$ ، تلاحظ أن $5.3 < 5.49$ إذن لا ترفض H_0. أنظر الشكل (14) أ، ب.





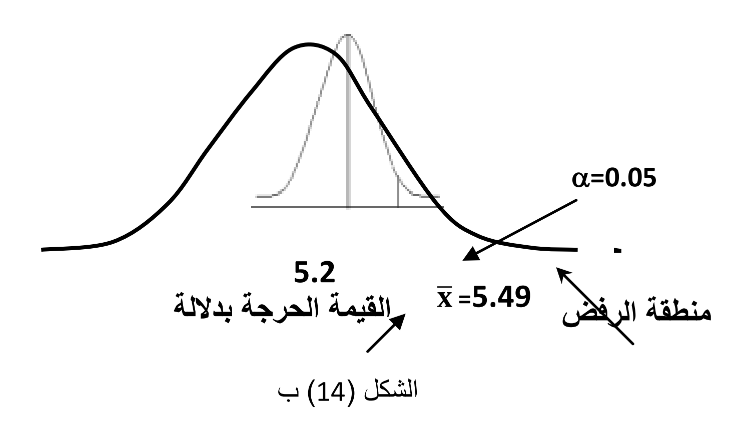
الشكل (14) أ

ب- الخطوات (1)-(4) تبقى كما هي.

(5) قيمة إحصاء الاختبار تحت فرصة H_0 صحيحة هي:

$$Z = \frac{5.3 - 5.2}{0.7/\sqrt{16}} = 0.56$$

(6) قارن قيمة Z = 0.56 مع المنطقة الحرجة، تجد 1.645 > 0.56|إذاً لا ترفض H_0. أنظر الشكل (14) f.



الشكل (14) ب

9 : 6 اختبار الفرضيات المتعلقة بالوسط لمجتمع طبيعي تباينه غير معلوم، وحجم العينة كبير.

إن عملية اختبار الفرضيات في هذه الحالة هي نفسها التي في (9 : 5) مع اختلاف واحد، وهو أنك تستعمل الانحراف المعياري للعينة S بدلا من الانحراف المعياري للتوزيع σ، وذلك لأن σ غير معلومة . أما تبرير إمكانية استعمال إحصاء الاختبار Z في هذه الحال فهو أن حجم العينة كبير أي $n \geq 30$.

مثال (10) :

تخضع أعداد حبات التفاح على شجرة التفاح في بستان كبير لتوزيع طبيعي وسطه 150 حبة. بدأ مالك البستان استعمال نوع جديد من السماد وأراد أن يختبر ما إذا زاد الإنتاج تبعا لذلك، لذا أخذ عينة من

64 شجرة، فوجد أن الوسط الحسابي لأعداد الحبات في العينة 156 بانحراف معياري 12 حبة. هل تشير هذه البيانات إلى الزيادة في الإنتاج على مستوى دلالة $\alpha = 0.05$ ؟

الحل : إذا لم يكن هناك زيادة في إنتاج الأشجار التي تم تسميدها فهذا يعني أن معدل عدد الحبات يكون $\mu = 150$، أما إذا كان هناك زيادة في الإنتاج فهذا يعني أن المعدل سيكون أكثر من 150.

فالمطلوب اختبار:

(1) $\mathbf{H_0 : \mu = 150}$

(2) $\mathbf{H_1 : \mu > 150}$

(3) $\boldsymbol{\alpha = 0.05}$

(4) إحصاء الاختبار : تحت فرض أن H_0 صحيحة فإن : $\mathbf{Z = \dfrac{\overline{X} - 150}{12/\sqrt{64}}}$ يخضع لتوزيع طبيعي معياري تقريبا.

(5) الفرضية البديلة ذات طرف واحد (أكبر من) إذن، القيمة الحرجة $Z_{0.95}$=1.645 والمنطقة الحرجة: ارفض H_0 إذا كان $Z > 1.645$.

(6) احسب Z من المعادلة :

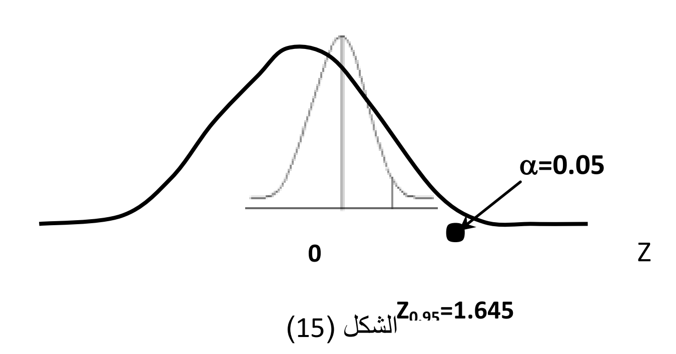
$$\mathbf{Z = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}}$$

أي

$$\mathbf{= \frac{156 - 150}{12/\sqrt{64}} = \frac{6}{12} \times 8 = 4}$$

(7) قارن قيمة Z التي حصلت عليها مع القيمة الحرجة :

من الواضح أن $4 > 1.645$ ، أي أن قيمة Z تقع في المنطقة الحرجة. لذلك ارفض H_0 لصالح $\mathbf{H_1 : \mu > 150}$. أنظر الشكل (15).



الشكل (15)

9 : 7 اختبار الفرضيات المتعلقة بالوسط لمجتمع تباينه معلوم وحجم العينة كبير :

Testing Hypotheses Concerning the Mean of a Population with Known Variance in Case large Sample Size.

إن عملية الاختبار في هذه الحالة هي نفسها التي في (9 : 5) إلا أن المجتمع في هذه الحالة لم يفترض أنه طبيعي، ولكن حجم العينة كبير $n \geq 30$. وبالتالي ، فإن المبرر لاستعمال إحصاء الاختبار Z هو نظرية تقارب التوزيعات والتي تضمن أن $Z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$ له توزيع قريب من التوزيع الطبيعي المعياري إذا كانت n كبيرة.

لاحظ أنه في حالة σ غير معلومة بإمكانك الاستعاضة عنها باستعمال S، الانحراف المعياري للعينة وذلك لأن حجم العينة كبير.

مثال (11) :

على مستوى 5% اختبر الفرضية $H_0 : \mu = 15$

مقابل الفرضية البديلة $H_1 : \mu \neq 15$

إذا أعطت عينة حجمها n = 81 وسطا حسابيا $\overline{X} = 13.5$ ، وانحرافا معياريا S = 3.3.

الحل :

(1) $H_0 : \mu = 15$

(2) $H_1 : \mu \neq 15$

(3) $a = 0.05$ مستوى الدلالة.

(4) إحصاء الاختبار : تحت فرض أن H_0 صحيحة فإن $Z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$ يخضع لتوزيع طبيعي معياري تقريبا.

(5) المنطقة الحرجة : الفرضية البديلة ذات طرفين، إذن ، فالقيم الحرجة هي :

$$Z_{0.025} = -1.96 \ , \ Z_{0.975} = 1.96$$

(6) الحسابات : احسب Z من المعادلة :

$$Z = \frac{\overline{X} - \mu_0}{S / \sqrt{n}}$$

$$= \frac{13.5 - 15}{3.3/\sqrt{81}} = \frac{-1.5 \times 9}{3.3} = -4.09$$

(7) المقارنة: قارن قيمة Z التي حصلت عيها مع القيم الحرجة : من الواضح أن 1.96- > 4.09-.

إذن ، ارفض H_0 لصالح $15 > \mu$. وقد قلت لصالح $15 > \mu$ لأن Z وقعت في المنطقة الحرجة إلى يسار القيمة الحرجة اليسرى.

9 : 8 اختبار الفرضيات المتعلقة بالوسط لمجتمع طبيعي تباينه غير معلوم وحجم العينة صغير.

لاحظ أن هذه الحالة تختلف عن تلك التي في (9 : 5) من حيث أن التباين σ^2 غير معلوم، وأنها تختلف عن الحالة في (9 : 7) من حيث حجم العينة صغير، وبالتالي لا تتحقق الشروط التي تبرر استعمال إحصاء الاختبار Z. بالرجوع إلى توزيعات المعاينة نجد أن توزيع الإحصاء :

$$T = \frac{\overline{X} - \mu_0}{S / \sqrt{n}}$$

هو توزيع t بدرجات حرية (n-1)، إذا كانت العينة التي حجمها n قد أخذت من توزيع طبيعي $N(\mu,\sigma^2)$ حيث $\overline{X}$ الوسط الحسابي ، S هو الانحراف المعياري لتلك العينة. إذن والحالة هذه، فإن إحصاء الاختبار المناسب لاختبار الفرضية حول μ هو :

$$T = \frac{\overline{X} - \mu_0}{S / \sqrt{n}}$$

ويتم إجراء الاختبار كما في (9 : 5) مع تعديل واحد، هو استعمال توزيع t بدرجات حرية (n-1) بدلا من استعمال التوزيع الطبيعي المعياري في إيجاد القيم الحرجة.

وخطوات الاختبار حول μ هي :

(1) $H_0 : \mu = \mu_0$

(2) مقابل الفرضية البديلة :

(i) $H_1 : \mu > \mu_0$

أو (ii) $H_1 : \mu > \mu_0$

أو (iii) $H_1 : \mu \neq \mu_0$

(3) مستوى الدلالة α.

(4) إحصاء الاختبار :

تحت فرض أن H_0 صحيحة، أي $\mu = \mu_0$ فإن $T = \frac{\overline{X} - \mu_0}{S / \sqrt{n}}$ تخضع لتوزيع t ذي درجات الحرية (n-1).

(5) النقاط الحرجة ومنطقة الرفض.

تتحدد النقاط الحرجة حسب الحالات الثلاث في الفرضية البديلة فتكون في الحالة :

(i) النقطة الحرجة $t_\alpha = -t[1-\alpha; n-1]$ ومنقطة الرفض هي : ارفض H_0 إذا كان t > t_α.

أما في الحالة (ii) فتكون :

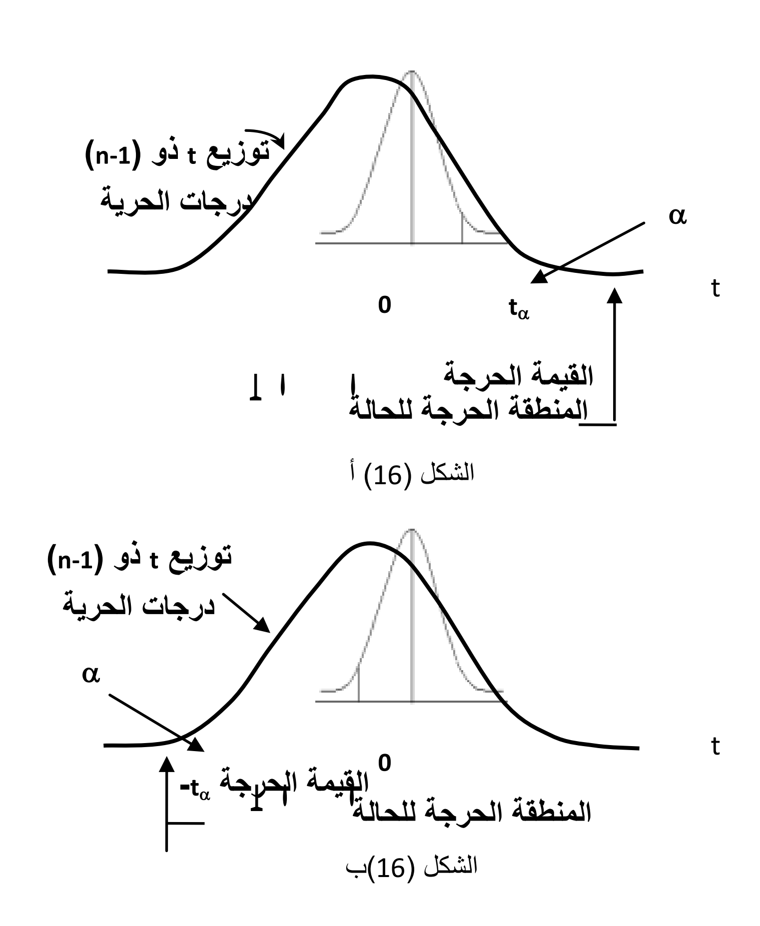
النقطة الحرجة هي $t_\alpha = t[1-\alpha; n-1]$ ومنطقة الرفض : أرفض H_0 إذا كان $t < -t_\alpha$.

أما في الحالة (iii) فتكون النقطتان الحرجتان $-t_{\frac{\alpha}{2}}\,,\,t_{\frac{\alpha}{2}}$ حيث $t_{\alpha/2} = t\left[1-\frac{\alpha}{2};n-1\right]$ ،

ومنطقة الرفض هي ارفض H_0 إذا كان $t < -t_{\frac{\alpha}{2}}$ أو $t > t_{\frac{\alpha}{2}}$ أي بعبارة أخرى إذا كان $|t| > t_{\frac{\alpha}{2}}$.

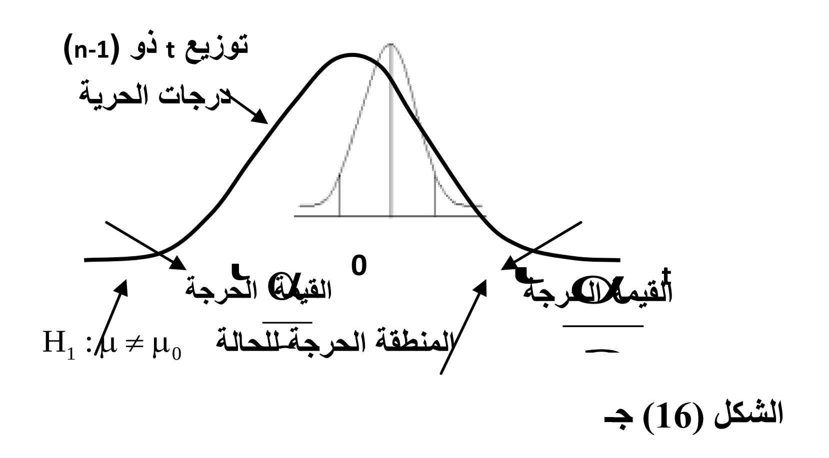
(6) احسب قيمة الاختبار T من (4) فتجد t.

(7) قارن t المحسوبة في (6) مع النقاط الحرجة في (5) واتخذ القرار المناسب، أي ارفض H_0 إذا وقعت t في منطقة الرفض كما في الأشكال (16) أ, ب, جـ.



الشكل (16) أ

الشكل (16)ب



الشكل (16) جـ

مثال (12) :

استنتج أحد الباحثين أن معدل عدد الساعات التي يقضيها طلبة إحدى الجامعات في الدراسة أثناء أسبوع الامتحانات 50 ساعة. اختبر هذه الفرضية مقابل فرضية أن معدل عدد الساعات يختلف عن 50 ساعة، إذا كان الوسط الحسابي لعدد الساعات التي قضاها 10 طلاب أثناء ذلك الأسبوع هو 51.7 ساعة بانحراف معياري 6.3 ساعة. استعمل مستوى دلالة 5% وافترض أن توزيع عدد الساعات الدراسية تقريبا طبيعي.

الحل :

1) الفرضية الصفرية $\mathbf{H_0 : \mu = 50}$

2) الفرضية البديلة $\mathbf{H_1 : \mu \neq 50}$

وقد تم استعمال الفرضية الثانية لأننا لا نعرف اتجاه الاختلاف أو الخطأ إذا كان استنتاج الباحث غير صحيح، ولذلك نأخذ الاتجاهين "أكبر من" "أصغر من" أي أن الفرضية البديلة هي $\mathbf{H_1 : \mu \neq 50}$.

3) مستوى الدلالة $\alpha = 0.05$.

4) بما أن الفرضية البديلة ذات اتجاهين فإن الاختبار المناسب ذو طرفين، وبما أن توزيع المجتمع طبيعي، تباينه غير معلوم، حجم العينة صغير، نستعمل توزيع t بدرجات حرية (n-1).

أي أن إحصاء الاختبار هو $\mathbf{t = \dfrac{\overline{X} - 50}{S / \sqrt{n}}}$.

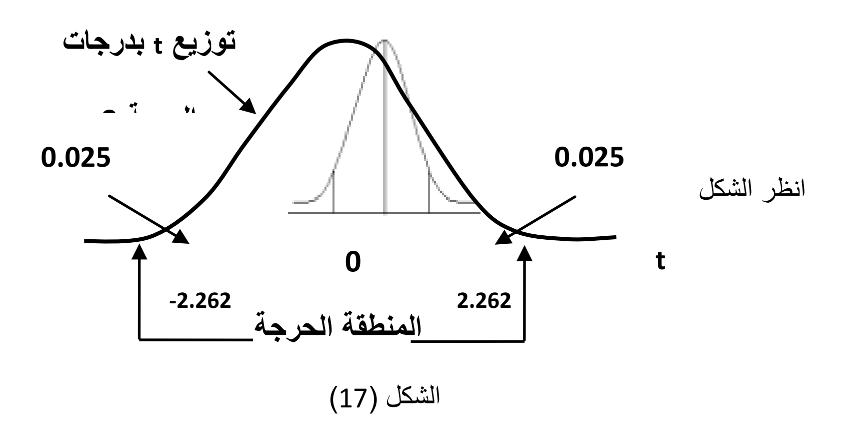
(5) بناء على (4) فإن :

القيم الحرجة هي $t_{0.025} = 2.262$ ، $-t_{0.025} = -2.262$

حيث درجات الحرية 9 = 1-10.

والمنطقة الحرجة: ارفض H_0 إذا كان|t| > 2.262.

انظر الشكل



الشكل (17)

(6) احسب t من المعادلة :

$$t = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

$$= \frac{51.7 - 50}{6.3/\sqrt{10}} = \frac{1.7}{6.3/3.16} = 0.85$$

(7) قارن قيمةt المحسوبة بالقيم الحرجة :

من الواضح أن 0.262 > 0.85 > 2.262-

إذن فإن t لا تقع في المنطقة الحرجة، وبالتالي فلا نرفض H_0.

مثال (13) :

كان معدل تحصيل طلبة احدى المدارس الخاصة في امتحان اللغة الإنجليزية الذي يتقدمون له عند طلب الالتحاق بالجامعات الأمريكية 400.

اختبر فرضية أن هذا المعدل قد تحسن إذا أعطت نتائج 14 طالبا وسطا حسابيا $\overline{X} = 408$ بانحراف معياري S = 23.

اعتبر أن نتائج طلبة المدارس تخضع لتوزيع طبيعي وخذ مستوى دلالة 1%.

الحل :

(1) $H_0 : \mu = 400$

(2) $H_1 : \mu > 400$

(3) مستوى الدلالة $\alpha = 0.01$.

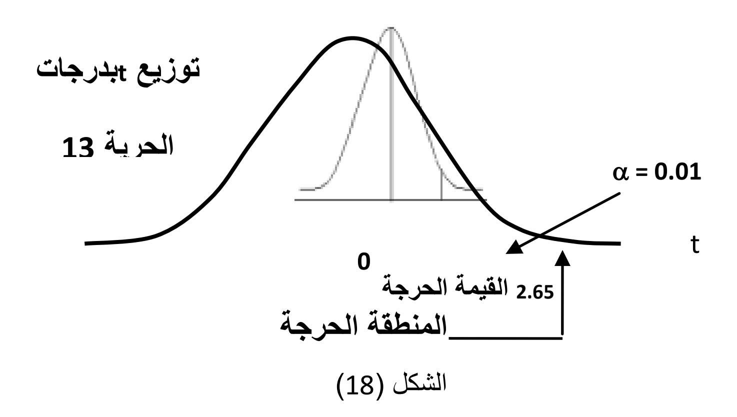
(4) إحصاء الاختبار هو $T = \dfrac{\overline{X} - 400}{S/\sqrt{n}}$ يخضع لتوزيع t ذي درجات حرية (n - 1).

(5) بما أن الفرضية البديلة لها اتجاه واحد فإن القيمة الحرجة $t_{0.01}$ تحت درجات حرية 13.

وقد تم استعمال توزيع t لأن المجتمع طبيعي ، تباينه غير معلوم، وحجم العينة 14 أي أقل من 30.

إذن القيمة الحرجة هي : $t_{0.01} = 2.65$

أنظر الشكل (18).



الشكل (18)

(6) احسب قيمة t من

$$t = \frac{\overline{X} - \mu_0}{S / \sqrt{n}}$$

$$= \frac{408 - 400}{23 / \sqrt{14}} = \frac{8 \times 3.742}{23}$$

$$= 1.3$$

(7) قارن بين قيمة t والقيمة الحرجة $t_{0.01}$.

من الواضح أن 2.65 > 1.3 أي أن t لا تقع في المنطقة الحرجة، وبالتالي لا نرفض H_0 أي لا نستطيع استنتاج أن معدل تحصيل الطلبة قد تحسن.

9 : 9 اختبار الفرضيات المتعلقة بالنسبة

Testing Hypotheses Concerning a Proportion

إن اختبار الفرضيات المتعلقة بالنسبة، أي نسبة المجتمع ذي خاصية معينة، يشبه اختبار الفرضيات المتعلقة بالوسط الحسابي، ويتلخص الأمر في تفحص، $\overline{P}$ ، النسبة في العينة التي تمتلك الخاصية المطلوبة، فإذا كانت هذه النسبة بعيدة جدا عن النسبة في المجتمع p_0 فإنك ترفض الفرضية الصفرية $H_0 : p=p_0$ وإلا فإنك ترفض H_1.

أما إحصاء الاختبار فهو **Z** حيث :

$$Z = \frac{\overline{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

وحيث توزيع **Z** قريب من التوزيع الطبيعي المعياري عندما تكون n كبيرة.

وخطوات إجراء الاختبار هي :

(أ) إذا أردنا النقاط الحرجة بدلالة Z تكون الخطوات :

(1) الفرضية الصفرية $H_0 : p = p_0$.

(2) مقابل الفرضية البديلة في احدى الحالات التالية :

(i) $H_1 : p > p_0$

$H_1 : \mathbf{p < p_0}$ (ii)

$H_1 : \mathbf{p \neq p_0}$ (iii)

(3) مستوى الدلالة α.

(4) إحصاء الاختبار : تحت فرض أن H_0 صحيحة وحجم العينة كبير يكون :

$$\mathbf{Z = \frac{\overline{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}}$$

يخضع لتوزيع طبيعي معياري تقريبا.

(5) النقاط الحرجة ومنطقة الرفض :

الحالة (i) النقطة الحرجة $z_{1-\alpha}$.

أرفض H_0 إذا كان $Z > z_{1-\alpha}$.

الحالة (ii) النقطة الحرجة z_α.

أرفض H_0 إذا كان $\mathbf{Z < z_\alpha}$.

الحالة (iii) النقطتان الحرجتان هما $\mathbf{z_{1-\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}}$.

أرفض H_0 إذا كان $\mathbf{Z < z_{\frac{\alpha}{2}}}$ أو $\mathbf{Z > z_{1-\frac{\alpha}{2}}}$.

وبعبارة أخرى نرفض H_0 إذا كان $|Z| > Z_{1-\frac{\alpha}{2}}$ لأن $Z_{\frac{\alpha}{2}} = - Z_{1-\frac{\alpha}{2}}$.

(6) نحسب **Z** من (4).

(7) نقارن قيمة **Z** التي حسبناها في (6) مع منطقة الرفض في (5) حسب حالة الاختبار ونقرر رفض H_0 أو عدمه على ضوء المقارنة ثم نذكر ذلك القرار بالكلمات.

(ب) إذا أردنا النقاط الحرجة بدلالة $\overline{P}$ ، فتبقى الخطوات كما هي ، ولكن في الخطوة (5) نحسب قيم $\overline{P}$ المقابلة للنقاط الحرجة ،

الحالة (i) : $$z_{1-\alpha} \equiv \frac{p_1 - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

إذا $$\overline{p}_1 = p_0 + z_{1-\alpha}\sqrt{\frac{p_0(1-p_0)}{n}}$$

ونرفض H_0 إذا كان $\overline{p} > \overline{p}_1$.

الحالة (ii) : $$\overline{p}_2 = p_0 + z_{\alpha}\sqrt{\frac{p_0(1-p_0)}{n}}$$

أو بعبارة أخرى :

$$\overline{p}_2 = p_0 - z_{1-\alpha}\sqrt{\frac{p_0(1-p_0)}{n}}$$

ونرفض H_0 إذا كان $\overline{P} < \overline{p}_2$

الحالة (iii) : $$\overline{p}_1 = p_0 + z_{1-\frac{\alpha}{2}}\sqrt{\frac{p_0(1-p_0)}{n}}$$

$$\overline{p}_2 = p_0 - z_{1-\frac{\alpha}{2}}\sqrt{\frac{p_0(1-p_0)}{n}}$$

ونرفض H_0 إذا كان $\overline{P} > \overline{p}_1$ أو $\overline{P} < \overline{p}_2$.

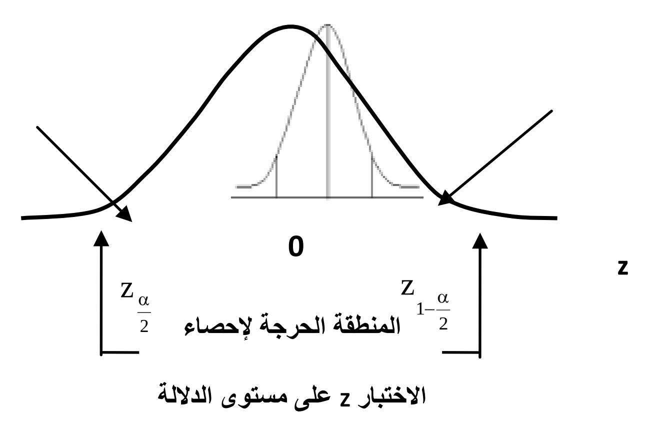
(6) لا نحتاج لهذه الخطوة لأن $\overline{P}$ محسوبة لدينا.

(7) المقارنة كما ذكرت في (7) سابقاً أي إذا وقعت $\overline{\mathbf{P}}$ في منطقة الرفض نرفض H_0 حسب الحالة المطلوبة ونذكر هذا القرار بالكلمات.

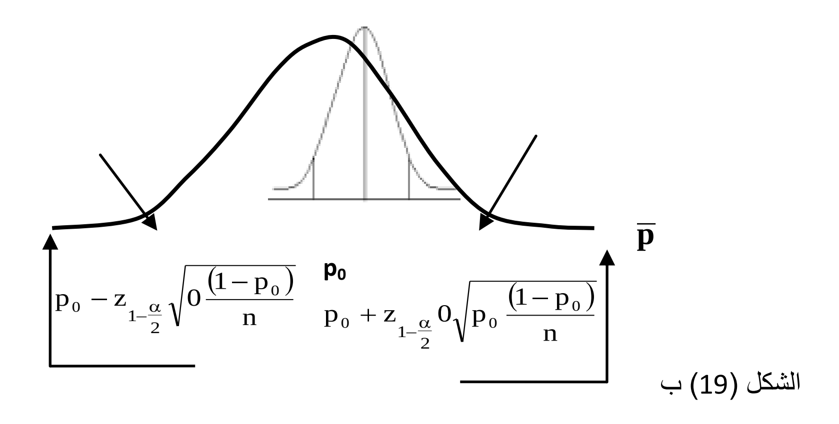
وكما تلاحظ ففي (أ) وجدنا النقاط الحرجة بدلالة Z من الجداول واحتجنا لحساب قيمة Z لكي نجري المقارنة. أما في (ب) فقد حسبنا القيم الحرجة المقابلة لقيم Z من الجدول واكتفينا بذلك فلا حاجة لحساب Z في الخطوة (6).

ومن ميزات (ب) أنه بعد إيجاد القيم الحرجة بدلالة $\overline{\mathbf{P}}$ فإننا نستطيع تطبيق الاختبار بدون حسابات إضافية لجميع العينات من ذات الحجم نفسه. ولكن إذا أردنا إجراء الاختبار مرة واحدة فيفضل إعطاء القيم الحرجة بدلالة Z.

أنظر الأشكال (19) أ،ب.



الشكل (19) أ



الشكل (19) ب

مثال (14) :

من المعلوم أن نسبة مستعملي حزام الأمان في السيارات (قبل تشريع الزام الاستعمال) هي 0.8. درست عينة عشوائية حجمها 200 سائق بعد صدور تشريع الالزام فوجد أن 170 منهم يستعلمون الحزام، اختبر على مستوى دلالة 5% ما إذا كان التشريع قد زاد نسبة المستعملين لحزام الأمان؟

الحل : من الواضح أن $\overline{P} = \frac{170}{200} = 0.85$

ضع خطوات الحل كما يلي :

(1) $H_0 : p = 0.80$ الفرضية الصفرية.

(2) $H_1 : p > 0.80$ الفرضية البديلة.

(3) $\alpha = 0.05$ مستوى الدلالة.

(4) إحصاء الاختبار : تحت فرض H_0 صحيحة فإن $Z = \frac{\overline{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ يخضع لتوزيع طبيعي معياري تقريبا.

(5) بما أن الفرضية البديلة ذات اتجاه واحد هو "أكبر من"، فإن الاختبار المناسب ذو طرف واحد، والقيمة الحرجة $z_{0.95} = 1.645$.

لاحظ إمكانية استعمال Z لأن حجم العينة كبير.

(6) احسب Z من المعادلة

$$Z = \frac{\overline{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

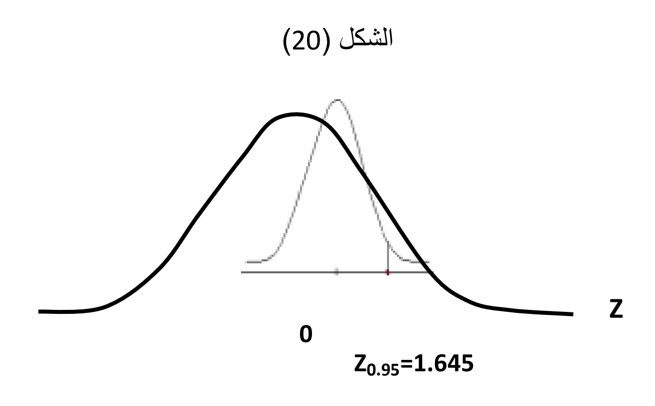
$$= \frac{0.85 - 0.80}{\sqrt{\frac{0.80 \times 0.20}{200}}} = \frac{0.05}{0.028} = 1.8$$

(7) قارن قيمة Z التي حصلت عليها مع القيمة الحرجة :

من الواضح أن $1.8 > 1.645$

إذن ارفض H_0 لصالح H_1 ، وبالتالي فإن صدور التشريع بالزامية استعمال حزام الأمان قد زاد من نسبة المستعملين له. شكل (20).

الشكل (20)



مثال (15) :

في المثال (14) ، ولدى دراسة عينات متعددة كل منها ذات حجم 200 في مدن مختلفة أعطت هذه العينات أعداد مستعملي حزام الأمان كما يلي :

المدينة (أ) 165 المدينة (ب) 172.

المدينة (جـ) 163 المدينة (د) 168.

ففي أي المدن كان التشريع قد زاد نسبة مستعملي حزام الأمان ؟

الحل : بما أننا سنجري الاختبار لعدة عينات من ذات الحجم نفسه فيفضل إعطاء القيم الحرجة بدلالة

 كما يلي :

(5) الخطوة السابقة تعطينا : النقطة الحرجة

$$\overline{P}_1 = p_0 + Z_{1-\alpha}\sqrt{\frac{p_0(1-p_0)}{n}}$$
$$= 0.8 + 1.645\sqrt{\frac{0.8\times 0.2}{200}}$$
$$= 0.8 + 0.046 = 0.846$$

ونرفض H_0 إذا كان $\overline{P} > 0.846$

المقارنة : نحسب $\overline{P}$ لكل مدينة ونقارن مع النقطة الحرجة فنجد :

المدينة	قيمة $\overline{P}$	القرار
أ	0.825	لا نرفض H_0
ب	0.86	نرفض H_0 لصالح H_1

لا نرفض H_0.	0.815	حـ
لا نرفض H_0.	0.84	د

9 : 10 اختبار الفرضيات المتعلقة بالفرق بين وسطين :

Testing Hypotheses Concerning the Difference Between Two means :

سنبحث هذا النوع من الاختبارات حسب كون المجتمعات طبيعية أم لا آخذين بعين الاعتبار معلومية التباينات أم عدمها وكبر حجوم العينات أو صغرها وتتلخص هذه بالنظريات التالية :

نظرية (2) :

إذا أخذت عينة عشوائية حجمها n_1 من مجتمع $N(\mu_1, \sigma_1^2)$ وأخذت عينة عشوائية ثانية ومستقلة عن الأولى حجمها n_2 من مجتمع $N(\mu_2, \sigma_2^2)$ وكان σ_2^2, σ_1^2 معلومتين فإن إحصاء الاختبار للفرضية $H_0 : \mu_1 = \mu_2$ هو :

$$Z = \frac{(\overline{X} - \overline{Y})}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$ الذي توزيعه N(0,1)

حيث $\overline{X}$ ، $\overline{Y}$ هما وسطا العينتين على التوالي.

من هذه النظرية نرى أن خطوات اختبار الفرضية $H_0 : \mu_1 = \mu_2$ هي نفسها الخطوات في السابقة مع الأخذ بعين الاعتبار إحصاء الاختبار الجديد.

أما إذا كانت المجتمعات لا تخضع للتوزيع الطبيعي فإننا نطبق نظرية تقارب التوزيعات (النهاية المركزية) إذا كانت حجوم العينات كبيرة.

نظرية (3) :

إذا أخذت عينتان عشوائيتان مستقلتان من مجتمعين معدل الأول μ_1 وتباينه σ_1^2 معلوم ومعدل الثاني μ_2 وتباينه σ_2^2 معلوم وكان حجم الأولى n_1 كبيرا وحجم الثانية n_2 كبيرا فإن إحصاء الاختبار للفرضية $H_0 : \mu_1 = \mu_2$ هو :

$$Z = \frac{\left(\overline{X} - \overline{Y}\right)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

الذي يخضع لتوزيع طبيعي معياري N(0,1) تقريبا.

إن خطوات اختبار الفرضية $H_0 : \mu_1 = \mu_2$ هي نفسها الخطوات السابقة مع الأخذ بعين الاعتبار إحصاء الاختبار الجديد.

ونلاحظ أيضا في هاتين النظريتين :

(1) إذا لم تكن التباينات معلومة فإننا نستطيع استعمال تباينات العينات S_2^2, S_1^2 ويبقى إحصاء الاختبار نفسه شريطة أن n_2, n_1 كبيرتان (كل منهما أكبر من 30).

(2) إذا كانت الفرضية الصفرية $H_0 : \mu_1 - \mu_2 = d$ حيث d عدد ثابت بدلا من الفرضية $H_0 : \mu_1 = \mu_2$ (أي $\mu_1 - \mu_2 = 0$) فإن إحصاء الاختبار في كلتي النظريتين يصبح :

$$Z = \frac{\left(\overline{X} - \overline{Y}\right) - d}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

مثال (16) :

أخذت عينتان مستقلتان حجماهما 72، 27 على التوالي من المجتمعين $N(\mu_1, 144)$، $N(\mu_2, 81)$ فأعطتا الوسطين $\overline{X} = 73$, $\overline{Y} = 69$.

أ) اختبر $H_0 : \mu_1 = \mu_2$ مقابل $H_1 : \mu_1 > \mu_2$ على مستوى دلالة $\alpha = 0.05$.

ب) اختبر الفرضية $H_0: \mu_1 - \mu_2 = 3$ مقابل $H_1: \mu_1 - \mu_2 > 3$ على مستوى دلالة $\alpha = 0.05$.

الحل : (أ)

(1) الفرضية الصفرية $H_0: \mu_1 = \mu_2$.

(2) الفرضية البديلة $H_1: \mu_1 > \mu_2$.

(3) مستوى الدلالة $\alpha = 0.05$.

(4) إحصاء الاختبار : تحت فرض أن H_0 صحيحة فإن إحصاء الاختبار هو :

$$Z = \frac{(\bar{X} - \bar{Y})}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

يخضع لتوزيع طبيعي معياري N (0,1).

(5) النقاط الحرجة ومنقطة الرفض: بما أن الاختبار ذو طرف واحد، فالنقطة الحرجة هي $Z_{0.95} =$ 1.645 ومنقطة الرفض : ارفض H_0 إذا كان $Z > 1.645$.

(6) الحسابات :

$$Z = \frac{73 - 69}{\sqrt{\frac{144}{72} + \frac{81}{27}}}$$

$$= \frac{4}{2.236} = 1.79$$

(7) أ- المقارنة : بما أن $1.79 > 1.645$ تقع في منطقة الرفض فإننا نرفض H_0 لصالح H_1.

ب- تبقى الخطوات الثلاث الأولى نفسها.

(4) إحصاء الاختبار : تحت فرض H_0 صحيحة (أي $H_0: \mu_1 - \mu_2 = 3$) فإن إحصاء الاختبار

$$Z = \frac{(\overline{X} - \overline{Y}) - 3}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

يخضع لتوزيع طبيعي N (0,1).

(5) منطقة الرفض تبقى نفسها. ارفض H_0 إذا كان Z > 1.645.

(6) الحسابات : $Z = \frac{(73 - 69) - 3}{\sqrt{\frac{144}{72} + \frac{81}{27}}}$

$$= \frac{1}{2.236} = 0.448$$

(7) المقارنة : من الواضح أن 1.645 $\not>$ 0.448

إذا لا نرفضH_0، أي أنه لا يوجد دلالة بأن الفرق بين الوسطين يزيد عن 3.

مثال (17) :

أخذت عينة عشوائية حجمها 100 من علامات الطالبات الناجحات في امتحان الثانوية العامة فأعطت وسطا حسابيا $\overline{X} = 68.5$ وانحرافاً معيارياً $S_1 = 10$.

وأخذت عينة عشوائية حجمها 150 من علامات الطلاب الذكور الناجحين في امتحان الثانوية العامة فأعطت وسطاً $\overline{Y} = 66.9$ وانحرافاً معيارياً $S_2 = 12$

اختبر على مستوى دلالة $\alpha = 0.05$.

أ) $H_0 : \mu_1 = \mu_2$ مقابل $H_1 : \mu_1 \neq \mu_2$.

ب) $H_0 : \mu_1 - \mu_2 = 1$ مقابل $H_1 : \mu_1 - \mu_2 \neq 1$.

على فرض أن μ_1 هو معدل الطالبات الناجحات ، μ_2 معدل الطلاب الذكور الناجحين في ذلك الامتحان.

الحل :

(أ) حجم كل عينة كبير (أكبر من 30) ولذلك شروط نظرية تقارب التوزيعات متحققة، ومع أن التباينات غير معلومة إلا أننا نقدرها بتباينات العينات لأن الحجوم كبيرة.

(1) $H_0 : \mu_1 = \mu_2$ مقابل الفرضية البديلة

(2) $H_1 : \mu_1 \neq \mu_2$.

(3) مستوى الدلالة $\alpha = 0.05$.

(4) إحصاء الاختبار : تحت فرض أن H_0 صحيحة وبالاستعاضة عن σ_2^2, σ_1^2 بتقديراتها S_2^2, S_1^2 على التوالي لأن حجوم العينات كبيرة فإن إحصاء الاختبار:

$$Z = \frac{\left(\overline{X} - \overline{Y}\right)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

يخضع لتوزيع طبيعي معياري تقريبا.

(5) منطقة الرفض :

بما أن الاختبار ذو طرفين فإن النقاط الحرجة هي :

$$Z_{1-\frac{\alpha}{2}} = Z_{0.975} = 1.96$$
$$Z_{\frac{\alpha}{2}} = Z_{0.025} = -1.96$$

والمنطقة الحرجة : أرفض H_0 إذا كان $|Z| > 1.96$

(6) الحسابات :

$$Z = \frac{68.5 - 66.9}{\sqrt{\frac{(10)^2}{100} + \frac{(12)^2}{150}}}$$

$$= \frac{1.6}{\sqrt{1 + 0.96}} = \frac{1.6}{1.4} = 1.14$$

(7) الاستنتاج : بما أن $1.96 \ngtr 1.14$.

إذاً لا نرفض H_0، أي لا يوجد دلالة بأن معدل الطالبات يختلف عن معدل الطلاب.

(ب) يبقى الحل كما هو ، إلا أن إحصاء الاختبار في (4) يصبح :

$$Z = \frac{(\bar{X} - \bar{Y}) - 1}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

والحسابات في (6) تصبح :

$$Z = \frac{(68.5 - 66.9) - 1}{\sqrt{\frac{(10)^2}{100} + \frac{(12)^2}{150}}}$$

$$= 0.43$$

الاستنتاج : بما أن $1.96 \ngtr 0.43$.

إذا لا نرفض H_0.

اختبار الفرضيات المتعلقة بالفرق بين وسطين عندما تكون التباينات غير معلومة وحجوم العينات صغيرة.

نظرية (4) :

أخذت العينة العشوائية التي حجمها n_1 من مجتمع طبيعي $N(\mu_1, \sigma_1^2)$ فأعطت الوسط الحسابي $\bar{X}$ والتباين S_1^2 وأخذت العينة العشوائية التي حجمها n_2 من مجتمع طبيعي $N(\mu_2, \sigma_2^2)$ مستقل عن

الأول فأعطت الوسط الحسابي $\overline{Y}$ والتباين S_2^2. فإذا كان $\sigma_1^2 = \sigma_2^2$ وكانت n_1, n_2 صغيرتين يكون إحصاء اختبار الفرضية $H_0 : \mu_1 = \mu_2$ هو :

$$T = \frac{(\overline{X} - \overline{Y})}{S_c\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

يخضع لتوزيع t ذي درجات الحرية $(n_1 + n_2 - 2)$ حيث :

$$S_c^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

لاحظ أنه بسبب صغر الحجمين n_1, n_2 فلا نستطيع استعمال نظرية النهاية المركزية.

مثال (18) :

أخذت عينة عشوائية حجمها 9 من أوزان الأطفال الذكور حديثي الولادة في أحد المستشفيات فأعطت $\overline{X} = 3.1$ ، $S_1 = 1.1$ (بالكغم) ، وأخذت عينة عشوائية حجمها 15 من أوزان الإناث حديثات الولادة في نفس المستشفى فأعطت $\overline{Y} = 2.8$ ، $S_2 = 1.3$ (بالكغم). على فرض أن كلا من أوزان الذكور وأوزان الإناث يخضع لتوزيع طبيعي ذي التباين نفسه، اختبر الفرضية :

(أ) $H_0 : \mu_1 = \mu_2$ مقابل $H_1 : \mu_1 > \mu_2$ على مستوى $\alpha = 0.05$.

(ب) $H_0 : \mu_1 - \mu_2 = 0.2$ مقابل $H_1 : \mu_1 - \mu_2 > 0.2$ على مستوى دلالة $\alpha = 0.05$ ، حيث μ_1 معدل أوزان الذكور، μ_2 معدل أوزان الإناث.

الحل :

(أ) شروط النظرية متحققة.

(1) $H_0 : \mu_1 = \mu_2$ مقابل الفرضية البديلة.

(2) $H_1 : \mu_1 > \mu_2$.

(3) مستوى الدلالة $\alpha = 0.05$.

(4) إحصاء الاختبار :

$$T = \frac{(\bar{X} - \bar{Y})}{S_c\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

يخضع لتوزيع t ذي درجات الحرية $n_1 + n_2 - 2$.

(5) منقطة الرفض : بما أن الاختبار ذو طرف واحد فإن النقطة الحرجة هي $t[1-\alpha; n_1 + n_2 - 2]$ ، $t[0.95;9+15-2] = 1.717$. ومنطقة الرفض : ارفض H_0 إذا كان $T > 1.717$.

(6) الحسابات : نحسب :

$$S_c^2 = \frac{8\times(1.1)^2 + 14\times(1.3)^2}{9+15-2} = \frac{32.46}{22} = 1.475$$

إذاً $S_c = \sqrt{1.475} = 1.21$

$$T = \frac{3.1 - 2.8}{1.21\sqrt{\frac{1}{9} + \frac{1}{15}}} = \frac{0.3}{1.21\sqrt{1.778}} = \frac{0.3}{0.51}$$
$$= 0.59$$

(7) المقارنة : بما أن $1.717 \not> 0.59$.

إذا لا نرفض H_0.

(ب) يبقى الحل نفسه إلا الخطوة (4) تصبح :

$$T = \frac{(\bar{X} - \bar{Y}) - 0.2}{S_c\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

والخطوة (6) تصبح :

$$T = \frac{(3.1 - 2.8) - 0.2}{1.21\sqrt{\frac{1}{9} + \frac{1}{15}}} = 0.197$$

(8) المقارنة : بما أن 1.717 $\ngtr$ 0.197.

إذا لا نرفض H_0 أي لا يوجد دلالة أن الفرق بين أوزان الأطفال الذكور حديثي الولادة وأوزان الأطفال الإناث حديثي الولادة يزيد على 0.2 كغم على مستوى 0.05.

9 : 11 المقارنــة بــين الأزواج المتقابلــة Matched Pairs Comparison :

يوجد في الحياة العملية تطبيقات على أزواج مرتبة ومتقابلة من البيانات، مثل الأزواج (X_i, Y_i)، i = 1, 2, ... , n حيث X_i صفة أو سمة للعنصر i Y_i، سمة أخرى للعنصر نفسه. ومن أهم الأمثلة

على ذلك إذا كانت X_i هي المشاهدات التي خضعت لمعاملة محددة (أ) وكانت Y_i المشاهدات التي خضعت لمعاملة أخرى (ب).

فلو فرضنا أن X_i حيث i = 1, ... , n عينة عشوائية من مجتمع معدله μ_x, وفرضنا $i = 1,...,n, Y_i$ عينة عشوائية من مجتمع معدله μ_Y فإننا نود المقارنة بين μ_Y, μ_X عن طريق فترات الثقة واختبار الفرضيات.

لقد قارنا بين معدلي مجتمعين بإيجاد فترة الثقة للفرق $\mu_1 - \mu_2$ وكذلك اختبرنا الفرضية $H_0 : \mu_1 = \mu_2$ تحت شرط أن العينتين مستقلتان، أي أننا أخذناهما من مجتمعين مستقلين، وهذا الفرض غير متحقق في حالة الأزواج المرتبة حيث أن الزوج (X_i, Y_i) مستقل عن الزوج (X_j, Y_j) لجميع $i \neq j$ ولكن X_i, Y_i ضمن الزوج الذي رقمه i غير مستقلة عن بعضها البعض لأن X_i مشاهدة عن العنصرi، Y_i مشاهدة عن العنصر i نفسه ولكن بعد إخضاع العنصر i لمعاملة جديدة.

ولإجراء المقارنة بين μ_Y, μ_X معدلي المجتمعين اللذين أخذنا منهما الأزواج (X_i, Y_i) ، i = 1, ... , n نرتب الخطوات كما يلي :

الفرق	السمة أو المعاملة (2)	السمة أو المعاملة (1)	رقم الزوج
$D_1 = X_1 - Y_1$	Y_1	X_1	1
$D_2 = X_2 - Y_2$	Y_2	X_2	2
$D_3 = X_3 - Y_3$	Y_3	X_3	3
.	.	.	.
.	.	.	.
.	.	.	.
$D_n = X_n - Y_n$	Y_n	X_n	n

تمثل الفروق $D_1, D_2, ..., D_n$ عينة عشوائية من مجتمع معدله δ وتباينه σ_D^2، فإذا افترضنا هذا المجتمع يخضع لتوزيع طبيعي $N(\delta, \sigma_D^2)$ أمكن إيجاد فترة ثقة $100(1-\alpha)\%$ للمعدل δ وأمكن اختبار الفرضية الصفرية $H_0 : \delta = \delta_0$ كما يلي:

احسب الوسط الحسابي والتباين للفروق :

$$\bar{D} = \frac{\sum D_i}{n}$$

احسب :

$$S_D^2 = \frac{\sum (D_i - \bar{D})^2}{n-1}$$

أي

$$S_D^2 = \frac{1}{n-1}\left[\sum D^2 - n\bar{D}^2\right]$$

فيكون :

(1) فترة ثقة $100(1-\alpha)\%$ للمعدل δ هي :

$$\left(\bar{D} - t_{\frac{\alpha}{2}} \frac{S_D}{\sqrt{n}}, \bar{D} + t_{\frac{\alpha}{2}} \frac{S_D}{\sqrt{n}}\right)$$

حيث $t_{\frac{\alpha}{2}} = t\left[1 - \frac{\alpha}{2}; n-1\right]$

(2) إحصاء الاختبار للفرضية $H_0 : \delta = \delta_0$ هو $T = \frac{\bar{D} - \delta_0}{S_D / \sqrt{n}}$ يخضع لتوزيع t ذي درجات الحرية (n-1).

مثال (19) : أخذت عينة عشوائية من 10 سكرتيرات وسجل عدد الكلمات التي تطبعها كل واحدة وليكن X_i.

تم تدريب السكرتيرات على الطباعة لمدة شهر وسجل عدد الكلمات التي تطبعها كل واحدة وليكن Y_i، فكانت النتيجة كما يلي :

الرقم	العدد قبل التدريب	العدد بعد التدريب	الفرق D_i	D_i^2
1	32	34	2	4
2	30	33	3	9
3	33	35	2	4
4	29	30	1	1
5	34	32	-2	4
6	36	34	-2	4
7	28	30	2	4
8	26	25	-1	1
9	31	29	-2	4
10	35	36	1	1

هل تدل هذه البيانات أن هناك فائدة للتدريب.

الحل : نحسب $\bar{D} = \frac{\sum D_i}{10}$ فنجد $\bar{D} = \frac{4}{10} = 0.4$

$$S_D^2 = \left[\sum D^2 - n\bar{D}^2\right] / n-1$$

$$= \frac{36 - 10 \times (0.4)^2}{9} = \frac{34.4}{9} = 3.82$$

$$S_D = 1.95$$

لو فرضنا أن μ_x معدل عدد كلمات الطباعة قبل التدريب و μ_Y المعدل بعد التدريب ، $\delta = \mu_Y - \mu_X$.

يكون المطلوب هو :

(1) اختبر الفرضية $H_0 : \delta = \delta_0 = 0$

(2) مقابل الفرضية البديلة $H_1 : \delta > 0$

(3) خذ مستوى الدلالة $\alpha = 0.05$.

(4) إحصاء الاختبار : تحت فرض H_0 صحيحة فإن : $T \equiv \frac{\overline{D}-0}{S_D/\sqrt{n}}$ يخضع لتوزيع t ذي درجات حرية (n-1).

(5) منطقة الرفض : بما أن الاختبار ذو طرف واحد $H_1 : \delta > 0$ ، فمنطقة الرفض هي : أرفض H_0 إذا كان $T > t[0.95;9] = 1.833$.

(6) الحسابات :

$$T = \frac{0.4}{1.95/\sqrt{10}} = 0.649$$

(7) المقارنة : بما أن $1.833 \ngtr 0.649$

فلا نرفض H_0 ، إذا لا يوجد دليل أن التدريب كان مفيدا.

مثال (20) :

يعطي الجدول التالي قياس ضغط الدم قبل أخذ دواء معين (X) وبعد تناول الدواء (Y) لخمس فتيات :

D^2	D = x - y	y بعد الدواء	x قبل الدواء
16	4	168	172
36	6	174	180
1	-1	173	172
4	-2	178	176
169	13	157	170

أ- أعط فترة ثقة 95% للمعدل δ.

ب- هل هناك دلالة أن الدواء خفض قياس ضغط الدم.

جـ- اختبر $H_0 : \delta = 0$ مقابل $H_1 : \delta \neq 0$.

$$\overline{D} = \frac{\sum Di}{n} = \frac{20}{5} = 4$$

$$S_D^2 = \frac{1}{4}\left[\sum d^2 - n\overline{d}^2\right]$$

$$= \frac{226 - 80}{4} = 36.5$$

$$S_D = 6.04$$

أ) فترة الثقة 95% للمعدل δ هي :

$$\left(\overline{D} - t_{0.025}\frac{S_D}{\sqrt{n}}, \overline{D} + t_{0.025}\frac{S_D}{\sqrt{n}}\right)$$

أي

$$4 \mp 2.776 \times \frac{6.04}{\sqrt{5}}$$

أي

$$(-3.5, 11.5)$$

(ب) المطلوب.

(1) اختبار $H_0 : \delta = 0$.

(2) مقابل الفرضية البديلة $H_1 : \delta > 0$ لاحظ أن انخفاض قياس ضغط الدم يعني $\mu_x > \mu_y$ أي $\mu_x - \mu_y > 0$.

(3) مستوى الدلالة $\alpha = 0.05$.

(4) إحصاء الاختبار تحت فرض H_0 صحيحة فإن $T = \frac{\overline{D}}{S_D / \sqrt{n}}$

يخضع لتوزيع t ذي (n-1) درجة حرية.

(5) منطقة الرفض : ارفض H_0 إذا كان $T > t[0.95;4] = 2.132$.

(6) الحسابات :

$$T = \frac{4}{6.04/\sqrt{5}} = 1.48$$

(7) المقارنة : بما أن $1.48 \ngtr 2.132$ فلا نرفض H_0.

(جـ) باتباع الخطوات كما في (ب) ولكن الآن منطقة الرفض هي:

$|T| > t\,[0.975, 4] = 2.776$

الحسابات كما هي أي $T = 1.48$

وبما أن $1.48 \ngtr 2.776$ فلا نرفض H_0.

لاحظ أ ن عدم رفض $H_0 : \delta = 0$ يقابله في فترات الثقة أن فترة الثقة تحوي الصفر، وبالرجوع إلى (أ) نجد أن فترة الثقة (11.5 ، 3.5-) تحوي الصفر أي أننا لا نرفض $H_0 : \delta = 0$

9 : 12 اختبار الفرضيات حول الفرق بين نسبتين

Testing Hypotheses Concerning the Difference Between Two Proportions :

للمقارنة بين نسبتين في مجتمعات بيرنوللي، أي ذات الحدين b (1, p)، بإمكاننا استعمال نظرية تقارب التوزيعات (النهاية المركزية) إذا كان حجم كل عينة كبيرا (أكبر من 30) حسب النظرية التالية :

نظرية (5) :

إذا أخذت عينة عشوائية حجمها n_1 من مجتمع $b\,(1, p_1)$ وأخذت عينة عشوائية ثانية مستقلة عن الأولى، وحجمها n_2 من مجتمع $b(1, p_2)$ فإن إحصاء الاختبار للفرضية $H_0 : p_1 = p_2$ هو:

$$Z = \frac{\overline{P}_1 - \overline{P}_2}{\sqrt{\dfrac{\overline{P}\left(1-\overline{P}\right)}{n_1} + \dfrac{\overline{P}\left(1-\overline{P}\right)}{n_2}}}$$

يخضع لتوزيع طبيعي معياري N (0,1) تقريبا إذا كانت n_2, n_1 كبيرتين ، حيث $\overline{P}_2, \overline{P}_1$ هي نسبة النجاح في العينة الأولى والعينة الثانية على التوالي ، $\overline{P}$ هي النسبة المشتركة ، أي

عدد النجاحات في العينة الأولى + عدد النجاحات في العينة الثانية
____
مجموع حجمي العينتين

$$\overline{P} = \frac{n_1\overline{P}_1 + n_2\overline{P}_2}{n_1 + n_2} =$$

ولإجراء الاختبار للفرضية $H_0 = p_1 = p_2$ نتبع الخطوات السابقة نفسها مع الأخذ بعين الاعتبار إحصاء الاختبار المعطى في النظرية (5).

مثال (21)

للمقارنة بين نسبة المدخنين في الفئة العمرية (18-25) سنة مع الفئة العمرية (26-30) سنة، أخذت عينة عشوائية حجمها 200 من الفئة الأولى فوجد أن 80 منهم يدخّنون، وأخذت عينة عشوائية مستقلة عن الأولى من الفئة العمرية الثانية وحجمها 100 فوجد أن 52 منهم مدخنون. اختبر الفرضية $H_0: p_1 = p_2$ مقابل $H_1: p_1 < p_2$ على مستوى دلالة $\alpha = 0.05$.

الحل:

(1) $H_0: p_1 = p_2$ مقابل الفرضية البديلة،

(2) $H_1: p_1 < p_2$.

(3) مستوى الدلالة $\alpha = 0.05$.

(4) إحصاء الاختبار هو

$$Z = \frac{\overline{P}_1 - \overline{P}_2}{\sqrt{\frac{\overline{P}(1-\overline{P})}{n_1} + \frac{\overline{P}(1-\overline{P})}{n_2}}}$$

يخضع لتوزيع (0, 1) N تقريبا لأن شروط النظرية متحققة حيث $n_1=200$ ، $n_2=100$ كبيرتان.

(5) منطقة الرفض.

بما أن الاختبار ذو طرف واحد $(H_1 : p_1 < p_2)$ فإن النقطة الحرجة هي $z_{0.05} = -1.645$.

ومنقطة الرفض : ارفض H_0 إذا كان $Z < -1.645$.

(6) الحسابات :

أولا نحسب :

$$\overline{P} = \frac{80+52}{200+100} = \frac{132}{300} = 0.44$$

$$\overline{P}_1 = \frac{80}{200} = 0.4 \ , \ \overline{P}_2 = \frac{52}{100} = 0.52$$

$$Z = \frac{-0.12}{\sqrt{0.0037}} = \frac{-0.12}{0.016} = -1.967$$

(7) المقارنة : بما أن $-1.967 < -1.645$ أي أن Z تقع في منطقة الرفض، إذاً نرفض H_0 لصالح H_1 أي أن نسبة المدخنين في الفئة العمرية الأولى أصغر من نسبة المدخنين في الفئة العمرية الثانية.

9 : 13 اختبار الفرضيات حول التباين

Testing Hypotheses Concerning Variance.

سنبحث اختبار الفرضيات حول التباين في حالة المعاينة من مجتمع طبيعي $N(\mu, \sigma^2)$ ويعطى إحصاء الاختبار كما في النظرية :

نظرية (6) :

إذا أخذت عينة عشوائية حجمها n من مجتمع طبيعي $N(\mu,\sigma^2)$ فإن إحصاء الاختبار للفرضية $H_0 : \sigma^2 = \sigma_0^2$ هو :

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

يخضع لتوزيع χ^2بدرجات حرية (n-1) حيث S^2 هو تباين العينة.

أما خطوات إجراء الاختبار فهي:

(1) $H_0 : \sigma^2 = \sigma_0^2$ مقابل الفرضية البديلة،

(2) الحالة (i) $H_1 : \sigma^2 < \sigma_0^2$

أو الحالة (ii) $H_1 : \sigma^2 > \sigma_0^2$

أو الحالة (iii) $H_1 : \sigma^2 \neq \sigma_0^2$

(3) مستوى الدلالة α.

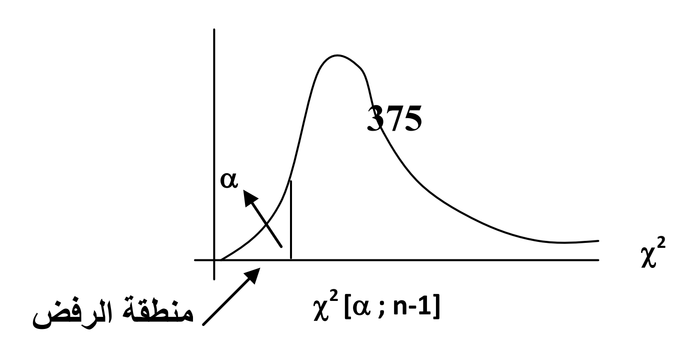
(4) إحصاء الاختبار : تحت فرض أن H_0 صحيحة (أي أن قيمة σ^2 هي σ_0^2 المعطاة في الفرضية) فإن إحصاء الاختبار

هو : $$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

يخضع لتوزيع χ^2 بدرجات حرية (n-1).

(5) النقاط الحرجة ومنطقة الرفض.

الحالة (i) النقطة الحرجة هي $\chi^2[\alpha;n-1]$ ومنطقة الرفض هي : ارفض H_0 إذا كان $\chi^2 < \chi^2[\alpha;n-1]$.

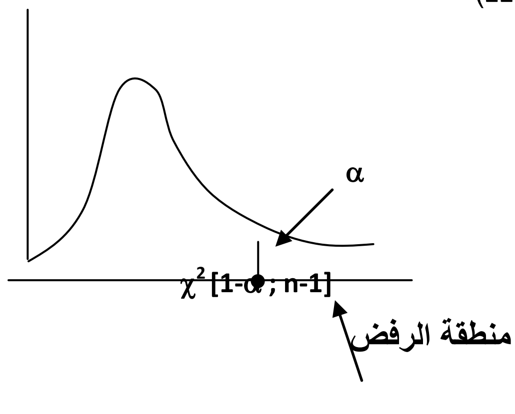


الشكل (21)

أنظر الشكل (21)

الحالة (ii) النقطة الحرجة هي $\chi^2[1-\alpha\,;\,n-1]$ ومنطقة الرفض هي ارفض H_0 إذا كان $\chi^2 > \chi^2\left[1-\alpha;n-1\right]$.

انظر الشكل (22)



الشكل (22)

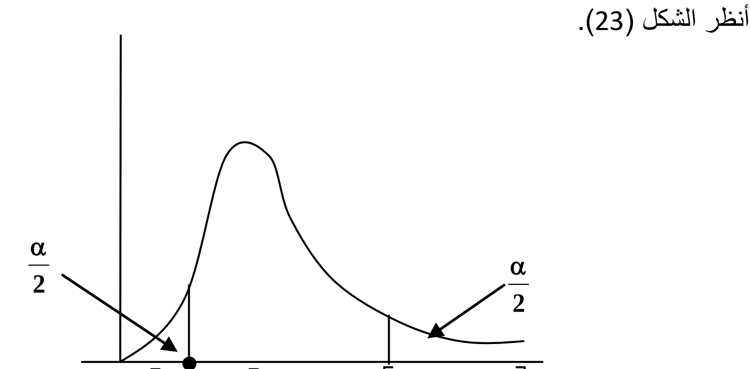
الحالة (iii) النقطتان الحرجتان هما

$$\chi^2\left[1-\frac{\alpha}{2};n-1\right] \;,\; \chi^2\left[\frac{\alpha}{2};n-1\right]$$

ومنطقة الرفض هي : أرفض H_0 إذا كان :

$$\chi^2 > \chi^2\left[1-\frac{\alpha}{2};n-1\right] \quad \text{أو} \quad \chi^2 < \chi^2\left[\frac{\alpha}{2};n-1\right]$$

أنظر الشكل (23).



الشكل (23)

(6) الحسابات : احسب قيمة χ^2 من (4).

(7) المقارنة : قارن فيما إذا كانت قيمة χ^2 المحسوبة في (6) تقع في منطقة الرفض وعندها ارفض H_0 وإذا كانت χ^2 المحسوبة لا تقع في منطقة الرفض فليس هناك دلالة على رفض الفرضية الصفرية H_0.

مثال (22) :

أخذت عينة عشوائية حجمها 9 من مجتمع $N(\mu,\sigma^2)$ فأعطت وسطاً حسابيا $\overline{X} = 56$ والتباين S^2 = 30.

اختبر على مستوى دلالة α = 0.05 الفرضية $H_0 : \sigma^2 = 33$ مقابل الفرضية $H_1 : \sigma^2 < 33$.

الحل :

(1) $H_0 : \sigma^2 = 33$ مقابل .

(2) $H_1 : \sigma^2 < 33$.

(3) مستوى الدلالة 0.05 = α.

(4) إحصاء الاختبار : تحت فرض أن H_0 صحيحة فإن $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$ يخضع لتوزيع ψ^2 ذي درجات الحرية (n-1).

(5) النقطة الحرجة، بما أن الاختبار ذو طرف واحد $(H_1 : \sigma^2 < 33)$ وأن n=9 فإن النقطة الحرجة هي $\chi^2[0.05;8]$.

ومنطقة الرفض هي : ارفض H_0 إذا كان $\chi^2 < 2.733$.

(6) الحسابات: نحسب ψ^2 من (4) حيث $\chi^2 = \frac{8 \times 30}{33} = 7.27$

(7) المقارنة : بما أن 2.733 $\nless$ 7.27 أي χ^2 لا تقع في منطقة الرفض، ولذلك فليس هناك دلالة على رفض $\sigma^2 = 33$.

مثال (23) :

في المثال (22) اختبر $H_0 : \sigma^2 = 33$ مقابل $H_1 : \sigma^2 > 33$ على المستوى $\alpha = 0.05$.

الحل :

نتبع خطوات حل المثال السابق نفسها إلا أن منطقة الرفض هي : ارفض H_0 إذا كان $\chi^2 > \chi^2[0.95;8] = 15.507$ وبما أن χ^2 المحسوبة كانت $\chi^2 = 7.27$ إذاً لا نرفض H_0 لأن 15.507 7.27.



9 : 14 المقارنة بين تبايني مجتمعين

Comparison Between two Variances

للمقارنة بين تبايني مجتمعين نختبر النسبة بين التباينين حسب النظرية:

نظرية (7) :

إذا أخذت عينة عشوائية حجمها n_1 من مجتمع $N(\mu_1, \sigma_1^2)$ وأخذت عينة مستقلة عن الأولى حجمها n_2 من مجتمع $N(\mu_2, \sigma_2^2)$ وكان S_1^2 تباين العينة الأولى، S_2^2 تباين العينة الثانية ، فإن إحصاء

الاختبار للفرضية $H_0 : \sigma_2^2 = \sigma_1^2$ هو S_2^2 / S_1^2 الذي يخضع تحت فرض H_0 صحيحة لتوزيع F ذي درجات الحرية (n_2-1) في البسط، (n_1-1) في المقام. أما خطوات إجراء الاختبار فهي :

(1) الفرضية الصفرية $H_0 : \sigma_1^2 = \sigma_2^2$ مقابل الفرضية البديلة :

(2) الحالة (i) $H_1 : \sigma_2^2 < \sigma_1^2$

الحالة (ii) $H_1 : \sigma_2^2 > \sigma_1^2$

الحالة (iii) $H_1 : \sigma_2^2 \neq \sigma_1^2$

(3) مستوى الدلالة α.

(4) إحصاء الاختبار : تحت فرض أن H_0 صحيحة فإن S_2^2 / S_1^2 يخضع لتوزيع $F[n_2 - 1, n_1 - 1]$.

(5) منطقة الرفض :

الحالة (i) ارفض H_0 إذا كان $S_2^2 / S_1^2 < F[\alpha; n_2 - 1, n_1 - 1]$.

الحالة (ii) ارفض H_0 إذا كان $S_2^2 / S_1^2 > F[1 - \alpha; n_2 - 1, n_1 - 1]$.

الحالة (iii) ارفض H_0 إذا كان S_2^2 / S_1^2 خارج الفترة :

$$\left(F\left[\frac{\alpha}{2}; n_2 - 1, n_1 - 1\right], F\left[1 - \frac{\alpha}{2}; n_2 - 1, n_1 - 1\right] \right)$$

(6) الحسابات : احسب S_2^2 / S_1^2.

(7) المقارنة : إذا وقعت S_2^2 / S_1^2 في منطقة الرفض حسب الحالة المعنية نرفض H_0.

مثال (24) :

أخذت عينتان مستقلتان من مجتمعين طبيعيين فأعطتا :

	العينة (1)	العينة (2)

الحجم	$n_1 = 8$	$n_2 = 10$
الوسط الحسابي	$\overline{X} = 27.4$	$\overline{Y} = 23.2$
التباين	$S_1^2 = 16$	$S_2^2 = 12$

اختبر على مستوى دلالة 0.025 الفرضية $H_0 : \sigma_2^2 = \sigma_1^2$ مقابل الفرضية $H_1 : \sigma_2^2 < \sigma_1^2$.

الحل :

(1) الفرضية الصفرية $H_0 : \sigma_2^2 = \sigma_1^2$.

(2) الفرضية البديلة $H_1 : \sigma_2^2 < \sigma_1^2$.

(3) مستوى الدلالة $\alpha = 0.05$.

(4) إحصاء الاختبار : تحت فرض أن H_0 صحيحة فإن S_2^2 / S_1^2 يخضع لتوزيع F [9 , 7].

(5) منطقة الرفض : بما أن الاختبار ذو طرف واحد وهي الحالة (i) نرفض H_0 إذا كان $S_2^2 / S_1^2 < F[0.05;9,7] = 0.238$.

(6) الحسابات : $\frac{S_2^2}{S_1^2} = \frac{12}{16} = 0.75$

(7) المقارنة : بما أن $0.75 \not< 0.238$.

إذا لا نرفض H_0 ، أي لا يوجد دلالة أن σ_1^2 أكبر من σ_2^2 .

ملاحظة : إن للنظرية السابقة أهمية في اختبار الفرضيات حول الفرق بين معدلي مجتمعين عندما تكون حجوم العينات صغيرة ولا يكون معلوما لدينا فيما إذا كان تباينا المجتمعين متساويين أم لا. وفي هذه الحالة يجب علينا أولا اختبار $H_0 : \sigma_1^2 = \sigma_2^2$ مقابل $H_1 : \sigma_1^2 \neq \sigma_2^2$ فإذا لم نرفض الفرضية H_0 كان لدينا تبرير بفرض تساوي التباينين وبعدها نستعمل إحصاء الاختبار T.

مثال (25) :

في المثال السابق إذا كان μ_1 معدل المجتمع الأول و μ_2 معدل المجتمع الثاني، اختبر

$H_0 : \mu_1 = \mu_2$ مقابل $H_1 : \mu_1 > \mu_2$ على مستوى دلالة $\alpha = 0.05$.

الحل : إن شروط نظرية استعمال إحصاء الاختبار T متحققة فيما عدا الفرض $\sigma_1^2 = \sigma_2^2$. ولذلك

نختبر أولا $H_0 : \sigma_1^2 = \sigma_2^2$ مقبل $H_1 : \sigma_1^2 \neq \sigma_2^2$ على مستوى $\alpha = 0.05$.

وذلك بإتباع خطوات الحل في المثال السابق وملاحظة أن منطقة الرفض هي خارج الفترة

$\left(F[0.025;9,7], F[0.975;9,7]\right)$ أي (0.238 , 4.82).

وبما أن $\frac{S_2^2}{S_1^2} = 0.75$ فهذا يعني أنها داخل الفترة (0.238 , 4.82) وبالتالي لا نرفض

$H_0 : \sigma_1^2 = \sigma_2^2$. إذاً الآن شروط نظرية إحصاء الاختبار T متحققة بما فيها شرط تساوي التباينين

ولذلك نتمكن من حل المسألة بالخطوات :

(1) $H_0 : \mu_1 = \mu_2$

(2) $H_1 : \mu_1 > \mu_2$.

(3) مستوى الدلالة $\alpha = 0.05$.

(4) إحصاء الاختبار : تحت فرض أن H_0 صحيحة فإن :

$$T = \frac{\overline{X} - \overline{Y}}{S_c \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

يخضع لتوزيع t ذي درجات الحرية $n_1 + n_2 - 2$.

(5) منطقة الرفض : $n_1 + n_2 - 2 = 8 + 10 - 2 = 16$

وبما أن الاختبار ذو طرف واحد، نرفض H_0 إذا كان $T > t[0.95,16] = 1.746$.

(6) الحسابات : $S_c^2 = \frac{7 \times 16 + 9 \times 12}{8 + 10 - 2}$

$$= \frac{220}{16}$$

$$S_c = 3.7$$

$$T = \frac{27.4 - 23.2}{3.7\sqrt{\frac{1}{8} + \frac{1}{10}}} = \frac{4.2}{1.755} = 2.39$$

(7) المقارنة : من الواضح أن $2.39 > 1.746$

إذا نرفض H_0 لصالح H_1.

9 : 15 اختبار الاستقلال Test of Independence

عندما ندرس نوعين من السمات أو الصفات لكل عنصر من عناصر عينة عشوائية فإن البيانات يمكن تصنيفها حسب السمتين أو الصفتين تحت الدراسة، وفي هذه الحالة نحصل على جدول توافق ثنائي التصنيف (Two-way contingency table) مجموع عناصر كل صف فيه عشوائي وكمثال على ذلك، سجل مدرس أحد المساقات عدد الطلبة لديه حسب الجنس والتخصص فوجد ما يلي :

		الجنس		
		ذكر	أنثى	المجموع
التخصص	رياضيات	25	7	32
	حاسب الكتروني	13	12	25
	إحصاء	22	11	33
	المجموع	60	30	90

إن مثل هذا الجدول يعرض بيانات نوعية Categorical data حيث أن المشاهدات توصف حسب تصنيفات وبالتالي تتألف البيانات من قيم التكرار في هذه التصنيفات. وتحدث مثل هذه البيانات بشكل كبير في معظم الدراسات الكمية وخاصة في الدراسات الاجتماعية والتربوية والسياسية.

إن أهم الاستدلال الإحصائي في جداول التصنيف الثنائي هذه هو معرفة فيما إذا كان التصنيفان أو السمتان اللتان استعملتا في التصنيف مستقلتين عن بعضهما أم غير مستقلتين وبعبارة أخرى هل

التصنيفان مستقلان عن بعضهما أم أن بعض مستويات أحد التصنيفين تميل أن ترتبط وتتوافق مع مستويات التصنيف الآخر.

فعلى سبيل المثال إذا صنف طلبة أحد المساقات حسب مدرسيهم أ ، ب ، جـ وحسب تقديرهم في المساق : ممتاز ، جيد جدا ، جيد ، مقبول ، راسب، فإن سؤالا هاما يطرح نفسه وهو هل التصنيفان (مدرس المساق، التقدير) مستقلان أم أن هناك توافق وارتباط مثلا بين تقدير ممتاز والمدرس ب أو هل هناك ارتباط وتوافق بين تقدير راسب والمدرس أ ؟ أي هل هناك علاقة بين المدرس وتقدير الطلبة ؟

وللإجابة على هذا السؤال نستعمل النظرية التالية :

نظرية (8) :

إذا صنفت مجموعة من المشاهدات وفق أسلوبين B , A في جدول توافق مدخلاته O_{ij} وأردنا اختبار الفرضية H_0 التي تنص على أن B , A مستقلان مقابل الفرضية H_1 القائلة بأنهما غير مستقلين فإننا نرفض H_0 على مستوى الدلالة α إذا كان :

$$U^2 = \sum_j \sum_i \frac{(O_{ij} - e_{ij})^2}{e_{ij}} > \chi^2[1-\alpha;(r-1)(c-1)]$$

حيث :

r عدد الصفوف ، c عدد الأعمدة ، O_{ij} عدد المشاهدات في الخلية (i , j). e_{ij} القيمة المتوقعة في الخلية (i , j) وقيمتها :

$$e_{ij} = \frac{(n_{i+})(n_{+j})}{n}$$

حيث : n_{i+} : مجموع المشاهدات في الصف الذي ترتيبه i.

n_{+j} : مجموع المشاهدات في العمود الذي ترتيبه j.

n: مجموع المشاهدات كلها.

مثال (26) :

أجريت دراسة لمعرفة آراء النـاس حـول تقديم التلفـاز لبرامج عنـف . صنفت عناصر العينـة حسب الجنس وحسـب الإجابـة عن السؤال : هل تـرى

أن هناك علاقة بين العنف في برامج التلفاز وانتشار الجريمة ؟

	الرأي			
	نعم	لا	غير متأكد	
ذكر	250	150	30	430
أنثى	300	50	20	370
المجموع	550	200	50	800

هل تدل هذه البيانات على فرق بين آراء الذكور والإناث ؟

الحل : للإجابة عن السؤال نعبر عنه بطريقة أخرى وهي : هل التصنيفان مستقلان، أي نختبر :

H_0 : التصنيفان مستقلان

H_1 : التصنيفان غير مستقلين.

مستوى الدلالة $\alpha = 0.05$.

نجري الحسابات التالية : نجد المجاميع $\mathbf{n_{i+}}$، $\mathbf{n_{+j}}$ ونحسب $\mathbf{e_{ij}}$ فنجد :

$$\mathbf{e_{11} = \frac{430 \times 550}{800} = 295.6}$$

$$\mathbf{e_{12} = \frac{430 \times 200}{800} = 107.5}$$

$$\mathbf{e_{13} = \frac{430 \times 50}{800} = 26.9}$$

$$e_{21} = \frac{370 \times 550}{800} = 254.4$$
$$e_{22} = \frac{370 \times 200}{800} = 29.5$$
$$e_{23} = \frac{370 \times 50}{800} = 23.1$$

ونحسب $U^2 = \sum\sum \frac{(O_{ij} - e_{ij})^2}{e_{ij}}$

$$= \frac{(250-295.6)^2}{295.6} + \frac{(150-107.5)^2}{107.5} + \frac{(30-26.9)^2}{26.9} + \frac{(300-254.4)^2}{254.4} + \frac{(50-92.5)^2}{92.5} + \frac{(20-23.1)^2}{23.1}$$
$$= 7.03 + 16.8 + 0.36 + 8.17 + 19.5 + 0.42$$
$$= 52.28$$

منطقة الرفض : أرفض H_0 إذا كان $U^2 > \chi^2[0.95;(2-1)(3-1)]$

$$\chi^2[0.95;2] = 5.99$$

المقارنة : بما أن 5.99 < 52.28 أي أن U^2 تقع في منطقة الرفض ولذلك نرفض H_0 أي أن التصنيفين غير مستقلين.

9 : 16 اختبار حسن المطابقة : Test of Goodness of Fit

نحتاج في كثير من الدراسات الإحصائية إلى التعرف فيما إذا كانت المشاهدات تتبع نموذجا معينا أو تخضع لتوزيع معين نفترضه، فعلى سبيل المثال إذا كان لدينا تكرارات القيم الظاهرة عند رمي زهرة نرد عددا كبيرا من المرات فهل نستطيع استنتاج أن الزهرة منتظمة ؟ وإذا سجلنا عدد الحوادث عند تقاطع طرق لعدد كبير من الأيام فإننا نود معرفة فيما إذا كان عدد الحوادث يخضع لتوزيع بواسون مثلا، وهكذا يكون لدينا مجموعة من المشاهدات موزعة على عدد من الخلايا ونرغب في اختبار فيما إذا كانت هذه المشاهدات تنطبق أو تحقق ما نتوقعه بفرض توزيع معين أو نسب معينة. للإجابة عن هذا التساؤل نطبق النظرية.

نظرية (9) :

إذا كان لدينا جدول تكراري لتجربة ما بحيث كان O_j تكرار الفئة E_j ، $j = 1, 2, \ldots, k$
ومجموع التكرارات N وأردنا اختبار الفرضية القائلة بأن المتغيرX الذي في هذا الجدول يخضع لنموذج إحصائي معين تعييناً كاملا فإننا نرفض هذه الفرضية إذا كان :

$$U^2 = \sum_{j=1} \left(O_j - e_j\right)^2 / e_j > \chi^2\left[1-\alpha; k-1\right]$$

حيث $e_j = NP\,(X \in E_j)$

وبشرط أن تكون N كبيرة كفاية لضمان $5 \geq e_j$ لكل j .

ملاحظات :

(1) قبل حل السؤال يجب التأكد أن $5 \geq e_j$ لكل j ، وإذا حدث أن هناك بعض الفئات التي يكون فيها التكرار المتوقع e_j أصغر من 5 فيجب دمجها مع أقرب فئة بجوارها حتى يتحقق الشرط $5 \geq e_j$.

(2) إذا كان في النموذج الإحصائي المقترح بعض المعلمات المجهولة، أي أن النموذج لا يتعين تعييناً كاملا إلا بعد معرفة قيم تلك المعلمات فإننا نقدر تلك المعلمات من العينة (الجدول التكراري الذي لدينا) وبعدها نطبق النظرية ودرجات الحرية في هذه الحالة = k - 1 - m حيث m هو عدد المعلمات التي تم تقديرها.

مثال (27) :

رميت زهرة نرد 120 مرة فكانت النتائج كما في الجدول التالي :

العدد الظاهر	1	2	3	4	5	6
التكرار	17	22	19	21	17	24

هل تدل هذه البيانات أن الزهرة منتظمة.

الحل: الزهرة منتظمة تعني $P(X = j) = \frac{1}{6}$ لجميع j=1, 2,...,6 وبما أنN= 120

إذا $e_j = 120 \times \frac{1}{6} = 20$ لجميع j=1, 2,...,6 إذاً شروط النظرية متحقق.

ونريد اختبار H_0 التي تنص أن الزهرة منتظمة أي احتمال ظهور أي عدد من 1 إلى 6 هو $\frac{1}{6}$.

$$U^2 = \sum_{j=1}^{6} \left(O_j - e_j\right)^2 / e_j$$

$$= \frac{(17-20)^2}{20} + \frac{(22-20)^2}{20} + \frac{(19-20)^2}{20} + \frac{(21-20)^2}{20} + \frac{(17-20)^2}{20} + \frac{(24-20)^2}{20}$$

$$= \frac{9}{20} + \frac{4}{20} + \frac{1}{20} + \frac{1}{20} + \frac{9}{20} + \frac{16}{20} = 2$$

منطقة الرفض : ارفض H_0 إذا كان

$$U^2 > \chi^2[1-\alpha; k-1] = \chi^2[0.95\,;5] = 11.07$$

وبما أن $11.07 \not> 2$ فلا نرفض الفرضية على مستوى دلالة $\alpha = 0.05$، أي نستنتج أن الزهرة منتظمة.

لاحظ أنه لو أردنا اختبار الفرضية H_0 على مستوى دلالة $\alpha = 0.10$ لكانت منطقة الرفض : ارفض H_0 إذا كان

$$U^2 > \chi^2[0.95\,;5] = 9.236$$

وبمقارنة $U^2 = 2$ مع القيمة الحرجة 9.236 فإن $9.236 \not> 2$ ولذلك لا نرفض H_0.

مثال (28) :

أخذت عينة عشوائية من 200 عائلة ذات الأربعة أطفال وسجل عدد الأبناء الذكور فكانت النتيجة كما في الجدول التالي :

عدد الذكور	عدد العائلات (التكرار O_j)

15	0
45	1
72	2
51	3
17	4

هل تدل هذه النتائج على أن عدد الأطفال الذكور في العائلة المكونة من 4 أبناء يخضع لتوزيع ذات الحدين $\mathbf{b}\left(4;\frac{1}{2}\right)$.

الحل : اختبر الفرضية H_0 القائلة أن X له التوزيع $\mathbf{b}\left(4;\frac{1}{2}\right)$. نلاحظ أن النموذج الإحصائي، أي توزيع X المقترح، أي $\mathbf{b}\left(4;\frac{1}{2}\right)$ معين تعيينا كاملا أي أنه لا يوجد معلمات تحتاج إلى التقدير، ونلاحظ عدد الفئات 5 ونحسب القيم المتوقعة e_j من احتمالات ذات الحدين المقترحة، فنجد :

$$\mathbf{e_1 = 200P(X=0) = 200 \times \binom{4}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^4 = 12.5}$$

$$\mathbf{e_2 = 200P(X=1) = 200 \times \binom{4}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^3 = 50}$$

$$\mathbf{e_3 = 200P(X=2) = 200 \times \binom{4}{2}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^2 = 75}$$

$$\mathbf{e_4 = 200P(X=3) = 200 \times \binom{4}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^1 = 50}$$

$$\mathbf{e_5 = 200P(X=4) = 200 \times \binom{4}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^0 = 12.5}$$

نحسب

$$\mathbf{U^2 = \sum \frac{(O_j - e_j)^2}{e_j}}$$

$$= \frac{(15 - 12.5)^2}{12.5} + \frac{(45 - 50)^2}{50} + \frac{(72 - 75)^2}{75} +$$
$$= \frac{(51 - 50)^2}{50} + \frac{(17 - 12.5)^2}{12.5}$$
$$= 0.5 + 0.5 + 0.12 + 0.02 + 1.62 = 2.76$$

منطقة الرفض : ارفض H_0 إذا كان

$$U^2 > \chi^2[0.95 ; 4] = 9.49$$

المقارنة : بما أن $9.49 \not< 2.76$ فلا نرفض H_0 ونستنتج أن العينة تعطينا أن النموذج الإحصائي المقترح صحيح.

مثال (29) :

في المثال السابق : هل يخضع عدد الذكور X لتوزيع ذات الحدين ؟

الحل : لاحظ أن عدد الأبناء في العائلة 4، فيصبح السؤال هل يخضع X لتوزيع $b\,(4\,;\,p)$ ؟

أي أن النموذج الإحصائي لم يتعين تماما ولذلك نقدر p بنسبة عدد الذكور للعائلة الواحدة في العينة $\overline{P}$ حيث :

$$\overline{P} = \frac{\sum x.O_j}{200 \times 4} = \frac{0 \times 15 + 1 \times 45 + 2 \times 72 + 3 \times 51 + 4 \times 71}{800} = \frac{1}{2}$$

نحسب القيم المتوقعة e_j (لاحظ أن $\overline{P} = \frac{1}{2}$ وهي نفسها القيمة في المثال السابق بمحض الصدفة)، ونجد أن e_j هي نفسها في المثال السابق، ونحسب U^2 فنجد $U^2 = 2.76$.

منطقة الرفض : ارفض H_0 إذا كان $U^2 > \chi^2[0.95 ; 3] = 7.81$

لاحظ أن درجات الحرية = (عدد الفئات - 1 - عدد المعلمات المقدرة) أي 5-1-1=3.

المقارنة : بما أن $7.81 \not< 2.76$ فلا نرفض H_0. لاحظ أنه لو كانت $\overline{P}$ التي حسبناها 0.4 مثلا لوجب علينا حساب e_j،$j = 1, 2, \ldots, 5$ من ذات الحدين $b\,(4; 0.4)$ وأكملنا الحل كالمعتاد.

تمارين

1-9 : إذا كان معدل نزول المطر على مدى السنوات السابقة في إحدى المحافظات هو 480 ملم وبانحراف معياري 40 ملم. أردت اختبار فيما إذا زاد معدل نزول المطر هذه السنة عن المعدل العام ولذلك أخذت عينة من 36 موقعا ووجدت أن المعدل 490 ملم وإذا قررت أن ترفض الفرضية إذا كان معدل العينة يبعد عن المعدل العام بأكثر من انحرافين معياريين للوسط الحسابي فأوجد المنطقة الحرجة واختبر الفرضية المذكورة.

2-9 : بناء على دراسات سابقة، يعتقد أن 0.35 من طلبة المدارس الثانوية مدخنون. قامت وزارة الصحة بحملة تثقيفية ضد التدخين لمدة عام، على أمل أن تؤدي إلى تقليل النسبة المذكورة، ولاختبار هذه المقولة، أخذت عينة حجمها 400 طالب من المدارس الثانوية فوجد أن عدد المدخنين 120 أوجد المنطقة الحرجة بدلالة $\overline{\mathbf{P}}$ (نسبة النجاح في العينة) وبدلالة Z وقم باختبار الفرضية إذا قررت أن ترفض الفرضية إذا كانت نسبة النجاح في العينة $\overline{\mathbf{P}}$ تبعد عن النسبة العامة 0.35 بأكثر من 1.64 انحراف معياري لنسبة النجاح.

3-9 : أجريت دراسات حول عدد ساعات تشغيل الطلبة في احدى الجامعات فوجد أن معدل عدد الساعات 16 ساعة والانحراف المعياري 7 ساعات.

ولمعرفة التغير في عدد ساعات عمل الطلبة قرر باحث اختبار الفرضية الصفرية :

$$\mathbf{H_0 : \mu = 16}$$

مقابل الفرضية البديلة $\mathbf{H_1 : \mu \neq 16}$

وقرر أن يرفض $\mathbf{H_0}$ إذا وقع $\overline{\mathbf{X}}$ (معدل العينة) بعيدا عن 16 بأكثر من انحرافين معياريين للوسط الحسابي.

احسب α إذا كان حجم العينة التي أخذها الباحث 50.

4-9 : أظهرت دراسات سابقة أن التباين في أعمار المصابيح الكهربائية التي ينتجها أحد المصانع يساوي (170) ساعة تربيع. اختبر الفرضية $\mathbf{H_0 : \mu = 850}$ مقابل الفرضية البديلة $\mathbf{H_1 : \mu \neq 850}$ إذا أعطت عينة عشوائية حجمها 100 مصباح وسطاً حسابيا $\overline{\mathbf{X}} = \mathbf{827}$. استعمل مستوى دلالة 5%.

9-5 : أخذت عينة عشوائية حجمها 10 من مجتمع طبيعي تباينه 150.

اختبر الفرضية $H_0 : \mu = 800$

مقابل الفرضية $H_1 : \mu \neq 800$ باستعمال مستوى معنوية 0.05 إذا كان الوسط الحسابي للعينة $\overline{X} = 812$.

9-6 : لمعرفة معدل عدد الساعات المعتمدة التي سجل لها طلبة احدى الجامعات في أحد الفصول ، أخذت عينة عشوائية حجمها 150 طالبا، فأظهرت العينة أن $\overline{X} = 12.5$ وأن S = 2.3 اختبر الفرضية.

$H_0 : \mu = 14$ مقابل الفرضية البديلة $H_1 : \mu < 14$ على مستوى دلالة $\alpha = 0.01$.

9-7 : وجد في دراسة سابقة أن معدل قيم الفواتير في أحد المستشفيات 70.2 دينارا وأن توزيعها يقرب من التوزيع الطبيعي .

أردت اختبار فرضية أن قيم الفواتير قد تغيرت فدرست 64 فاتورة أخذت عشوائياً فوجد أن $\overline{X} = 73.7$ وأن S = 11.2.

باستعمال مستوى دلالة 5% اختبر الفرضية $H_0 : \mu = 70.2$ مقابل الفرضية البديلة المناسبة.

9-8 : يبلغ معدل طول الجندي في أحد الجيوش 169سم وفي السنوات الأخيرة بدأ الإقبال على الجندية يزيد وبالتالي صارت القيادة تضع شروطا أشد بخصوص الطول. اختبر الفرضية التي تقول إن معدل طول الجندي في ذلك الجيش قد ازداد، علما بأن عينة عشوائية حجمها 50 من أفراد ذلك الجيش أعطت معدل الطول 171.5سم والانحراف المعياري 5سم. استعمل $\alpha = 0.05$.

9-9 : إذا كانت نسبة العائلات التي تملك البيوت التي تسكن فيها في مدينة معينة هي 0.7 أجريت إحصائية عن 2000 موظف فوجد أن 1480 شخص من بينهم يملكون البيوت التي يسكنونها. هل نستطيع أن نستنتج أن نسبة الموظفين المالكين للبيوت التي يسكنونها أعلى من النسبة العامة ؟

9-10 : على مستوى دلالة 0.05 .

(أ) أوجد القيمة الحرجة بدلالة Z لاختبار الفرضية $\mathbf{H_0 : \mu = 50}$ مقابل الفرضية البديلة $\mathbf{H_1 : \mu < 50}$. حيث μ هو وسط المجتمع، ومفترضا أن حجم العينة كبير بشكل كاف ليسمح باستعمال التوزيع الطبيعي وإحصاء الاختبار Z.

(ب) إذا كان S = 9 وحجم العينة n = 64 ، $\overline{\mathbf{X}}$ هو الوسط الحسابي للعينة ، أوجد المنطقة الحرجة والقيمة الحرجة بدلالة $\overline{\mathbf{X}}$.

9-11 : اختبرت 50 عينة من مياه أحد السدود فوجد أن معدل ذوبان مكونات الأكسجين هو 13.7 جزء لكل مليون وانحراف معياري 3.2 جزءا لكل مليون. هل تعطي هذه البيانات دليلا كافيا على أن تركيز الأكسجين أقل من 15.5 جزء لكل مليون ؟ استعمل $\mathbf{\alpha = 0.05}$.

9-12 : أعطت دراسة سابقة استنتاجا بأن معدل الزمن الذي تحتاجه العاملات في معمل خياطة لإنجاز عدد من القطع الجاهزة هو 7 ساعات.

اختبر هذا الاستنتاج مقابل فرضية أن المعدل أقل من 7 ساعات إذا أعطت عينة عشوائية حجمها 23 عاملة $\overline{\mathbf{X}} = \mathbf{6.5}$ ساعة وانحراف معياري S = 1.4 ساعة.

افرض أن الزمن موزع حسب التوزيع الطبيعي واستعمل مستوى دلالة 0.05.

9-13 : اختبر الفرضية $\mathbf{H_0 : \mu = 12}$

مقابل الفرضية البديلة $\mathbf{H_1 : \mu < 12}$ على مستوى 1%.

إذا أعطت عينة عشوائية حجمها 8 من مجتمع طبيعي النتائج التالية

$\mathbf{S = 2.7, \overline{X} = 6.5}$.

9-14 : في تمرين (8-8) ، اختبر الفرضية $\mathbf{H_0 : \mu = 38}$ مقابل $\mathbf{H_1 : \mu > 38}$ ، على مستوى دلالة 0.05.

9-15 : في تمرين (8-14)، اختبر الفرضية $\mathbf{H_0 : \mu = 17}$ مقابل $\mathbf{H_1 : \mu > 17}$ ، على مستوى دلالة 0.05.

واختبر الفرضية $H_0 : \sigma^2 = 4$ مقابل $H_1 : \sigma^2 < 4$، على مستوى دلالة 0.05.

16-9 : في تمرين(8-18) اختبر الفرضية $H_0 : \sigma^2 = 15$ مقابل $H_1 : \sigma^2 \neq 15$، على مستوى دلالة، 0.05.

17-9 : في تمرين (8-22) اختبر الفرضية $H_0 : \mu_1 = \mu_2$ مقابل الفرضية $H_1 : \mu_1 > \mu_2$ على مستوى دلالة 0.05 حيث μ_1 المعدل في البحيرة (أ) μ_2 المعدل في البحيرة (ب).

18-9 : في تمرين (8-23)

(أ) اختبر الفرضية $H_0 : \sigma_1^2 = \sigma_2^2$ مقابل $H_1: \sigma^2_1 \neq \sigma^2_2$ على مستوى دلالة $\alpha = 0.05$.

(ب) اختبر الفرضية $H_0 : \mu_1 = \mu_2$ مقابل $H_1 : \mu_1 > \mu_2$ على مستوى دلالة 0.05.

هل تحتاج لنتيجة (أ) لتحل (ب) ؟

19-9 : في التمرين (8-21) اختبر فرضية أن المعدلين الحقيقيين متساويان.

20-9 : في التمرين (8-25) اختبر الفرضية : $H_0 : \mu_1 - \mu_2 = 5$ مقابل $H_1 : \mu_1 - \mu_2 > 5$ على مستوى دلالة 0.05.

21-9 : في التمرين (8-26) اختبر الفرضية $H_0 : \mu_1 - \mu_2 = 3$ مقابل $H_0 : \mu_1 - \mu_2 > 3$ على مستوى دلالة 0.05.

22-9 : يبين الجدول التالي النبض لعدد من الطلبة قبل الركض مائة متر (X) وبعد الركض (Y).

Y	X
75	70
74	72
68	70

73	71
75	72
78	76

هل تظهر هذه البيانات أن النبض يزيد بعد الركض. استعمل $\alpha = 0.01$.

23-9 : أخذت عينة عشوائية وصنفت حسب المؤهل العلمي والرأي حول الصوت الواحد في الانتخابات فكانت النتيجة كما يلي :

	موافق على الصوت الواحد	غير موافق	متردد
ثانوية عامة فما دون	70	35	55
بكالوريوس	40	80	30
دراسات عليا بعد البكالوريوس	30	55	15

هل تدل هذه البيانات على أن المؤهل العلمي والرأي حول الصوت الواحد مستقلان على مستوى دلالة 0.05.

24-9 : للمقارنة بين تأثير دواءين أ, ب ضد السعال في زيادة النوم أخذت عينة عشوائية من 8 مرضى وتناولوا الدواء (أ) في الليلة الأولى والدواء (ب) في الليلة الثانية وسجلت ساعات نومهم في كل ليلة فكانت :

	المريض							
	1	2	3	4	5	6	7	8
الدواء (أ)	3.8	4.5	7.7	6.8	7.2	6.7	5.3	7.4
الدواء (ب)	4.6	5.2	7.1	6.5	7.3	7.0	5.5	7.1

أ. أوجد فترة ثقة 95% لمعدل التغير في عدد ساعات النوم عند الانتقال من الدواء (أ) إلى الدواء (ب).

ب. اختبر على مستوى 0.05 فيما إذا كان هناك فرق بين ساعات النوم لدى استعمال الدواءين في ليلتين مختلفتين.

25-9 : يمثل الجدول التالي التوزيع التكراري لعدد الأخطاء المطبعية لكل صفحة من كتاب مكون من 200 صفحة.

عدد الأخطاء المطبعية X	عدد الصفحات O_j
0	97
1	45
2	27
3	23
4	8

هل يخضع عدد الأخطاء X لتوزيع بواسون ؟ استعمل $\alpha = 0.05$.

26-9 : قابل طبيبان نفسانيان أ ، ب عددا من المراجعين وسجلا فيما إذا كان المريض يعاني من انفصام الشخصية أم لا وحصلا على التصنيف التالي :

	المرض موجود	المرض غير موجود
تشخيص (أ)	25	5
تشخيص (ب)	22	8

هل تدل البيانات على وجود فرق جوهري بين آراء الطبيبين ؟

27-9 : أخذت عينة عشوائية حجمها 100 من مجتمع طبيعي معدله μ_1 وتباينه σ_1^2 وأخذت عينة عشوائية مستقلة عن الأولى حجمها 80 من مجتمع طبيعي معدله μ_2 وتباينه σ_2^2 فأعطتا ملخص الإحصاءات التالية:

	العينة الأولى	العينة الثانية
الوسط الحسابي	$\overline{X} = 67.2$	$\overline{Y} = 70.8$
الانحراف المعياري	$S_1 = 11$	$S_2 = 12$

أ- اختبر $H_0 : \sigma_1^2 = 100$ مقابل $H_1 : \sigma_1^2 > 100$.

ب- اختبر $H_0 : \mu_2 = 70$ مقابل $H_1 : \mu_2 > 70$.

جـ- اختبر $H_0 : \mu_1 = \mu_2$ مقابل $H_1 : \mu_1 < \mu_2$.

د- اختبر $H_0 : \sigma_1^2 = \sigma_2^2$ مقابل $H_1 : \sigma_2^2 > \sigma_1^2$.

هـ- أوجد فترة ثقة 95% للانحراف المعياري σ_1. استعمل $\alpha = 0.05$ في جميع الفروع.

الفصل العاشر
الارتباط والانحدار
Correlation and Regression

الارتباط Correlation

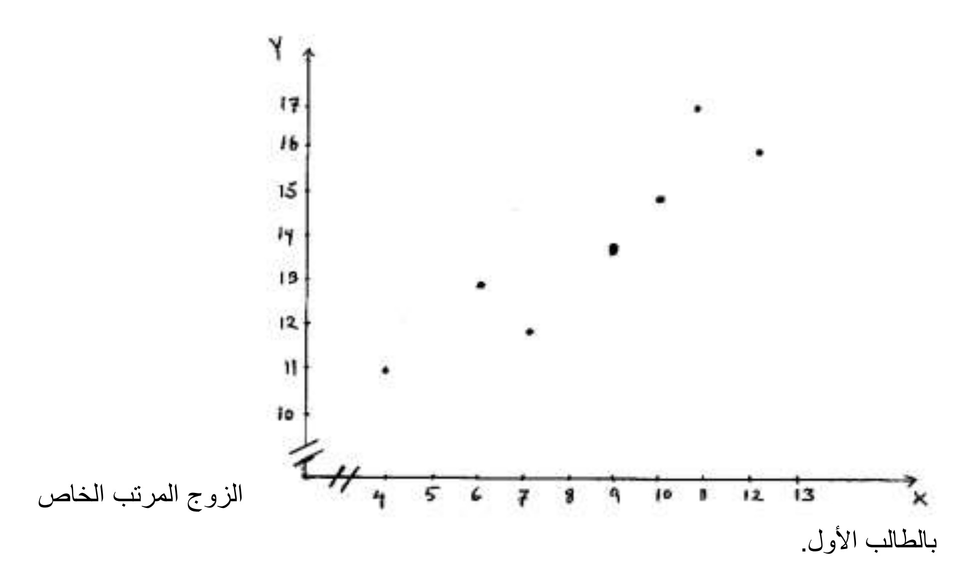
1-10 : مقدمة Introduction

لقد بحثنا في الفصول السابقة بعض المسائل المتعلقة بقياسات ومشاهدات عن متغير واحد، وبعبارة أخرى انحصرت دراستنا في موضوع ظاهرة واحدة، أي أن البيانات التي كانت لدينا كانت قياسات أو مشاهدات من نوع واحد تم تسجيلها عن مجموعة من الأفراد أو العناصر. فمثلا : كانت البيانات عن الدخل الشهري لمجموعة من الموظفين، أو علامات مجموعة من الطلبة في أحد الامتحانات، أو أوزان الأطفال حديثي الولادة وهكذا. ونلاحظ أن دراستنا كانت عن وضع هذه البيانات في توزيع تكراري، أو حساب مقاييس النزعة المركزية لها، أو مقاييس التشتت وفي مثل هذه الحالات تكون دراستنا عن متغير واحد أو متغير ذي البعد الواحد.

أما في هذا الفصل فندرس قياسين عن كل عنصر من العناصر قيد الدرس، كأن نسجل طول ووزن كل طالب في احدى المدارس الأساسية ومن ثم ندرس العلاقة بين هذين القياسين أو المتغيرين. وفي كثير من الأحيان نعبر عن المتغيرين عبارة "متغير ذو بعدين".

والأمثلة على ذلك كثيرة ، فقد تحتاج إلى معرفة العلاقة بين التحصيل في الامتحان النهائي لدى الطالب في فصل معين وبين عدد الساعات التي درسها الطالب في تحضيره لذلك الامتحان.

في هذه الحال نسجل قيمتين أو مشاهدتين عن كل طالب، الأولى عدد ساعات الدراسة والثانية علامة الطالب في الامتحان النهائي. وبذلك نكون قد سجلنا زوجا مرتبا من القيم عن كل طالب. فإذا عبّرنا عن عدد ساعات الدراسة بالحرف X وعلامة الطالب بالحرف Y نحصل على ما يلي :



الزوج المرتب الخاص بالطالب الأول.

الزوج المرتب الخاص بالطالب الثاني.

وهكذا حتى نصل إلى الذي يمثل الزوج المرتب الخاص بالطالب الذي رقمه n.

ويمكن وضع مثل هذه القياسات على شكل جدول، كما يظهر في الجدول(1).

جدول (1)

x	6	13	19	8	10	3	7	1	15
y	67	75	86	61	71	50	62	40	80

وهذا يعني ، أن عدد ساعات دراسة الطالب الأول كان 6 ساعات وعلامته 67، أما الطالب الثاني فدراسته 13 ساعة وعلامته 75، فيما نجد أن دراسة ا لطالب السادس يساوي 3 ساعات وعلامته 50.

ندرس في هذا الفصل العلاقة بين متغيرين، نسمي أحدهما X والآخرY ، ونضع القيم لهذين المتغيرين إما في جدول أو على شكل أزواج مرتبة.

فإذا كان لدينا n من الأزواج المرتبة.

فهذا يعني أن لدينا عينة عشوائية من قيم المتغيرين X و Y.

وبدراسة هذه العينة من الأزواج المرتبة (x, y) نريد الإجابة عن السؤالين :

1- هل هناك علاقة بين المتغيرين ؟ وهل هي خطية أم غير خطية ؟ أي هل نستطيع أن نحكم أن النقاط (x, y) تقع على خط مستقيم أم لا تقع ؟

وبعبارة أخرى ، هل يرتبط أحد المتغيرين بالآخر، كأن يزيد أحدهما مع الازدياد في الآخر، أو ينقص أحدهما إذا ازداد الآخر ؟ إن معرفة وجود علاقة بين المتغيرين وقياس قوة تلك العلاقة هي موضوع الارتباط، وإذا كانت العلاقة خطية فإن المقياس الذي نقيس به قوة العلاقة الخطية هو معامل الارتباط الخطي.

فمثلا ، هل هناك علاقة بين ضغط الدم والوزن لدى الرجال ؟ إذا كانت هناك علاقة نقول إن هناك ارتباطا بينهما. أما إذا كانت العلاقة خطية أي إذا كان يصاحب الزيادة المعينة في الوزن زيادة محددة في الضغط تقريبا فنقول هناك ارتباط خطي بين المتغيرين.

2- والآن إذا كان هناك علاقة بين المتغيرين فكيف نعبر عنها بمعادلة ؟

إن معرفة المعادلة التي تربط بين المتغيرين تفيد في تنبؤ قيمة أحدهما إذا علمت قيمة الآخر، فلو عرفت المعادلة التي تربط بين علامة الطالب في شهادة الدراسة الثانوية ومعدله عند التخرج في الجامعة لاستطعنا تقدير معدل طالب ما عند التخرج في الجامعة، أو التنبؤ به، إذا علمنا علامته في شهادة الدراسة الثانوية.

ويسمى المتغير الأول (هنا العلامة في شهادة الدراسة الثانوية) المتغير المستقل، فيما يسمى المتغير الثاني (معدل الطالب عند التخرج) المتغير التابع.

ومن الأمثلة في الاقتصاد ، معرفة العلاقة بين دخل الفرد الشهري وتوفيره، ومعرفة العلاقة بين تكاليف الدعاية وحجم المبيعات من سلعة معينة .

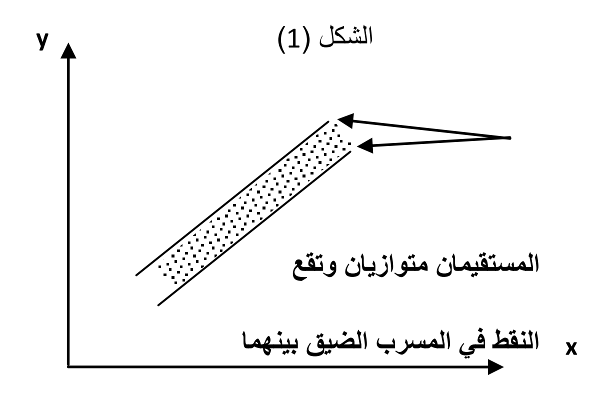
10 : 2 لوحة الانتشار Scatter Diagram

إن إحدى الطرق التي تساعدنا على معرفة وجود علاقة أو علاقة خطية بين متغيرين Y , X هي لوحة الانتشار حيث نرسم إحداثيا أفقيا (إحداثي X) وإحداثيا عموديا عليه (إحداثي Y).

ونرصد أزواج المشاهدات المرتبة :

المعطاة لدينا على المستوى XY فنحصل على لوحة الانتشار.

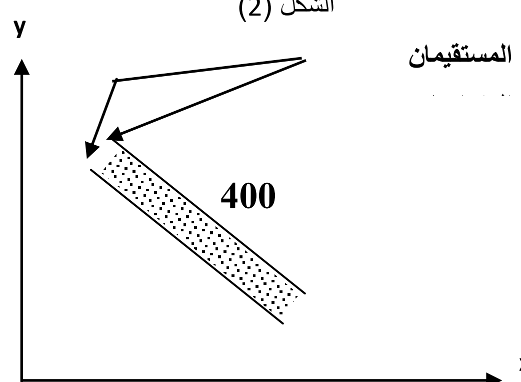
وتبين لك لوحة الانتشار بشكل جيد ما إذا كان هناك علاقة بين المتغيرين Y, X أو عدم وجودها. وإضافة إلى ذلك فإن لوحة الانتشار تبين ما إذا كانت النقاط (x, y) واقعة على خط مستقيم و في مسرب ضيق بين مستقيمين متوازيين. وهذا يوحي إمكانية وقوعها على خط مستقيم بمعنى إمكانية رسم خط مستقيم تكون معظم النقط واقعة عليه أو قريبة منه، أو أن هذه النقط متبعثرة بشكل يتضح منه عدم وقوعها على خط مستقيم. ويوضح الشكل (1) أن النقط في لوحة الانتشار تقع في مسرب ضيق، بين مستقيمين متوازيين، ولذلك يتبين إمكانية وقوع هذه النقط على خط مستقيم، بمعنى أنه يمكنك رسم خط مستقيم بحيث تقع معظم النقط عليه أو تكون حوله وقريبة منه.



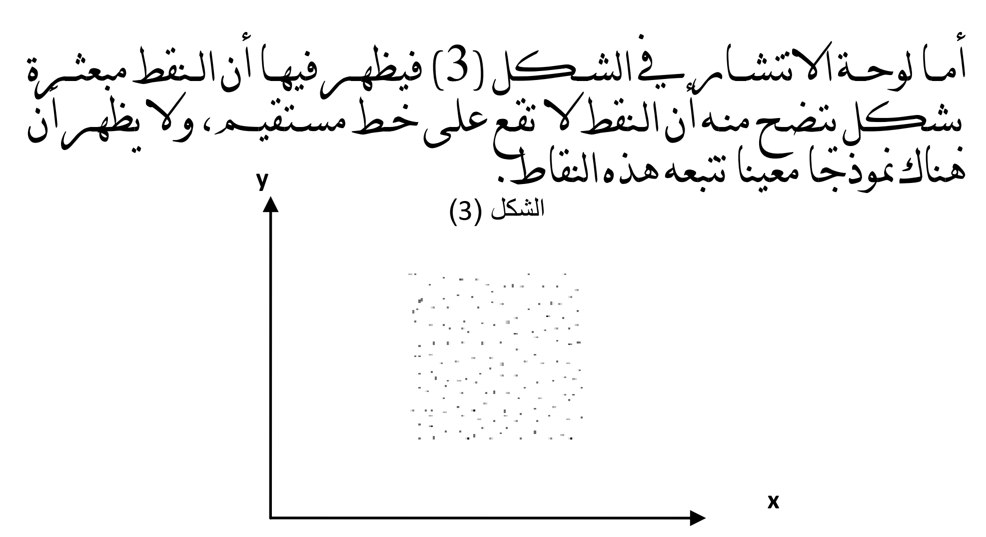
الشكل (1)

وكذلك الحال في شكل (2) ، تقع النقاط التي رصدتها والتي تمثل الأزواج المرتبة في مسرب ضيق بين خطين مستقيمين متوازيين. وبذلك تتبين مكانية وقوع هذه النقاط على خط مستقيم بمعنى إمكانية رسم خط مستقيم تقع معظم النقاط عليه أو تكون قريبة منه. وبالتالي إمكانية وجود علاقة خطية بين لمتغيرين Y, X .

الشكل (2)



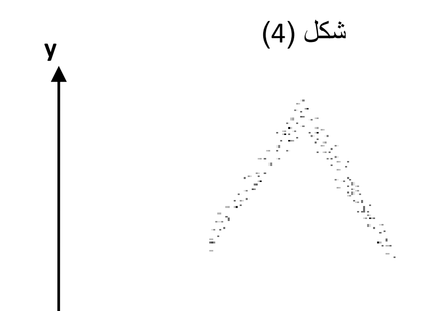
أما لوحة الانتشار في الشكل (3) فيظهر فيها أن النقط مبعثرة بشكل يتضح منه أن النقط لا تقع على خط مستقيم، ولا يظهر أن هناك نموذجا معينا تتبعه هذه النقاط.



الشكل (3)

أما لوحة الانتشار في الشكل (4) فيظهر منها أن هناك علاقة بين المتغيرين X, Y كما ويظهر أن هذه العلاقة ليست علاقة خطية حيث النقاط المرصودة لا توحي بأنها تقع على خط مستقيم بل على منحنى اقتران تربيعي.

شكل (4)



إضافة إلى ما سبق فإن لوحة الانتشار تساعد في توضيح كون العلاقة بين المتغيرين طردية أم عكسية. فإذا نظرنا إلى النقاط التي تم رصدها على لوحة الانتشار في الشكل (1) نستنتج أنه كلما ازدادت القيم x ازدادت القيم y، أي كلما سرنا إلى اليمين على المحور الأفقي نجد أن قيم y ترتفع إلى أعلى.

إذن فالعلاقة هنا علاقة طردية أي أن القيم y تزداد مع ازدياد القيم x وتنقص مع نقصانها.

أما الشكل (2) فيظهر فيه أن القيم y تنقص مع ازدياد القيم x والعكس بالعكس، وهذا يعني أن العلاقة بين المتغيرين عكسية.

مثال (1) :
الجدول (2) يعطي معدلات عشر طلاب في شهادة الدراسة الثانوية (X) ومعدلاتهم في نهاية الفصل الأول في إحدى الجامعات (Y).

معدل الطالب في شهادة الدراسة الثانوية X	معدل الطالب في نهاية الفصل الأول Y
77.5	61.0
85.5	72.8
79.9	65.5
88.8	71.0
93.5	80.0
85.0	74.0
94.1	82.0
90.5	81.0
87.2	76.1
89.8	78.0

الجدول (2)

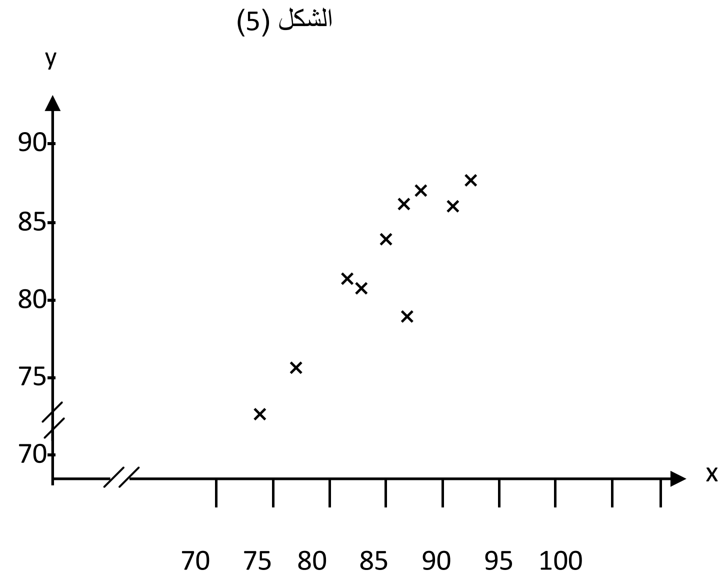
ارسم لوحة الانتشار. ما نوع العلاقة بين معدل الطالب في شهادة الدراسة الثانوية ومعدله في الفصل الأول الجامعي ؟

الحل : افرض X تمثل معدل الطالب في شهادة الدراسة الثانوية . وافرض Y تمثل معدل الطالب في الفصل الجامعي.

ارسم خطين متعامدين يمثل الأول (الأفقي)محورx ويمثل العمودي عليه محورy.

ارصد على المستوى xy النقط التي احداثياتها الأزواج المرتبة من الجدول أي النقاط : , (77.5, 61.0) (89.8, 78.0) , ... , (85.5, 72.8)كما يظهر في الشكل (5).

الشكل (5)



ولاحظ أننا تركنا في بداية المحور x فراغا وفي بداية المحور y ليدل على أن المسافة بين المركز والقيمة 70 محور على x وكذلك المسافة بين المركز والقيمة 60 على محور y قد قطعتا لكي تتمكن من استعمال مقياس رسم مناسب.

لاحظ أن لوحة الانتشار تظهر بوضوح أن العلاقة بين القيم y والقيم x هي علاقة خطية ، وهي أيضا علاقة طردية، لأن معدل الطالب في الفصل الجامعي الأول يرتفع بارتفاع معدله في شهادة الثانوية العامة.

مثال (2) :سجل أحد المعلمين عدد الساعات التي قضاها سبعة طلاب تم اختيارهم عشوائيا من الصف التاسع الأساسي في التحضير لامتحان الرياضيات وعلامـاتهم في ذلك الامتحان في أول الأسبوع التالي. فوجدها كما في الجدول (3).

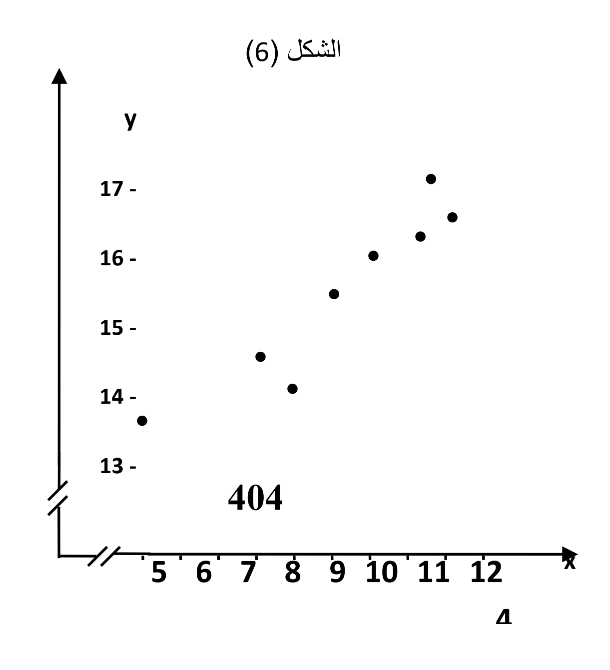
عدد الساعات التي قضاها الطالب للتحضير للامتحان (x)	علامة الطالب في امتحان الرياضيات (y)
6	13
12	16
4	11
11	17
10	15
7	12
9	14

الجدول (3)

ارسم لوحة الانتشار وأذكر العلاقة بين علامة امتحان الرياضيات وعدد ساعات الدراسة.

الحل : افرض x يمثل عدد ساعات الدراسة . وافرض y يمثل علامة الطالب في امتحان الرياضيات.
على المستوى xy ، ارصد النقط التي احداثياتها الأزواج المرتبة : ,9) , ... , (18 , 12) , (6, 13) .10)

الشكل (6)



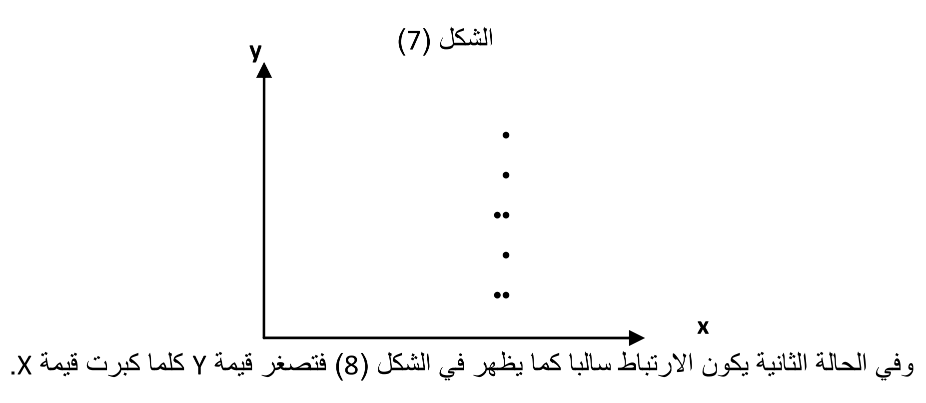
يظهر من لوحة الانتشار أنه يمكن اعتبار العلاقة بين عدد ساعات الدراسة والعلامة في امتحان الرياضيات علاقة خطية، وكما يظهر فهي علاقة طردية أي كلما ازدادت X ازدادت Y.

10 : 3 معامل الارتباط الخطي : Coefficient of Linear Correlation

إن دراسة الارتباط تظهر إلى أي حد يمكن أن يتحرك متغيران معا، وذلك بإعطاء مقياس عددي يحدد درجة تلك الحركة.

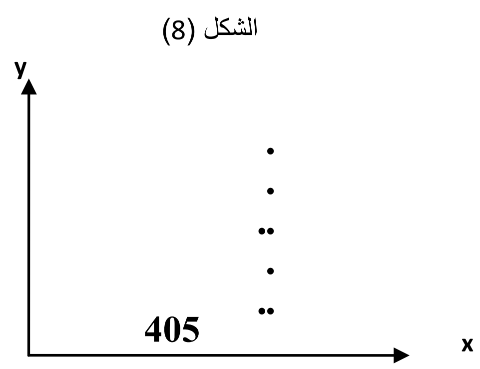
ومعامل الارتباط الخطي هو مقياس قوة العلاقة الخطية بين المتغيرين. وهو يقيس مقدار التغير والتأثير الذي يطرأ على Y عندما يزداد X مقدارا معينا. إنه يعطي فكرة فيما إذا كانتY تزداد كلما ازدادت X أو أنها تنقص كلما ازدادت X أو أنها لا تتبع نمطا محددا في الزيادة والنقصان. في الحالة الأولى يكون الارتباط موجبا، أي تزداد Y , X معا فكلما كبرت X كبرت Y. كما يظهر في الشكل (7).

الشكل (7)



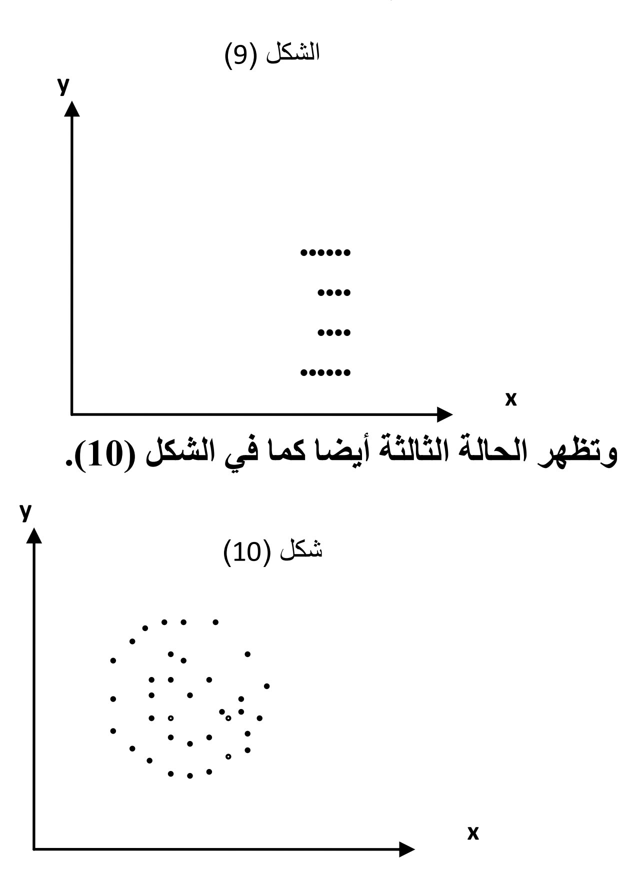
وفي الحالة الثانية يكون الارتباط سالبا كما يظهر في الشكل (8) فتصغر قيمة Y كلما كبرت قيمة X.

الشكل (8)



أما الحالة الثالثة فكما هو واضح في الشكل (9) فإن Y فيها لا تتبع نمطا محددا عند ازدياد X وفي هذه الحال نقول إنه لا يوجد ارتباط بين X, Y.

الشكل (9)



وتظهر الحالة الثالثة أيضا كما في الشكل (10).

شكل (10)

تعريف (1) : معامل ارتباط بيرسون الخطي لمجموعة n من الأزواج المرتبة ونعبر عنه بالرمز r هو مجموع حواصل ضرب القيم المعيارية للقيم x_i مع القيم المعيارية المقابلة لها للقيم y_i مقسوما على (n-1) أي :

حيث : هو الوسط الحسابي للقيم

$\bar{y}$ هو الوسط الحسابي للقيم $y_1, y_2, ..., y_n$

s_x هو الانحراف المعياري للقيم $x_1, x_2, ..., x_n$

s_y هو الانحراف المعياري للقيم $y_1, y_2, ..., y_n$

إن حساب قيمة r من المعادلة السابقة يحتاج إلى وقت ، وخاصة إذا كانت الأوساط الحسابية تحتوي كسورا. لذلك نكتب تعريف r على شكل صالح للاستعمال بالآلات الحاسبة وهو :

$$r = \frac{\sum xy - n\bar{x}\bar{y}}{\sqrt{\sum x^2 - n\bar{x}^2}\sqrt{\sum y^2 - n\bar{y}^2}}$$

وهذه الصيغة استعملت سابقا في حساب التباين والتباين المشترك. لاحظ أننا لتسهيل الكتابة توقفنا عن كتابة الرمز i مع كل x أو y أو في رمز الجمع $\sum$، أي أننا من الآن فصاعدا وحيثما لا يكون هناك التباس نكتب $\sum x$ بدلا من $\sum_{i=1}^{n} x_i$.

ولإيجاد قيمة معامل ارتباط بيرسون الخطي لمجموعة من الأزواج المرتبة من البيانات نستعمل المعادلة في التعريف إذا كانت البيانات صغيرة العدد أو طلب منا حساب القيم المعيارية للبيانات إضافة إلى حساب معامل الارتباط.

أما في معظم الحالات فإننا ننصح باستعمال معادلة الآلات الحاسبة لسهولتها وسهولة استعمال الآلة الحاسبة اليدوية في إجراء العمليات الحسابية الداخلة فيها.

ولكي تجري الحسابات المطلوبة لإيجاد قيمة معامل الارتباط بين متغيرين نقوم بترتيب العمليات الحسابية كما في الأمثلة التالية :

مثال (3) :

احسب معامل الارتباط بين المتغيرين Y , X حيث تكون قيمة كل من X , Yكما في الجدول (4) التالي :

X	8	1	6	4	7	7	2
Y	6	4	3	4	5	4	2

الحل : رتب خطوات الحل كما في الجدول (5) التالي :

x	y	xy	x^2	y^2
8	6	48	64	36
1	4	4	1	16
6	3	18	36	9
4	4	16	16	16
7	5	35	49	25
7	4	28	49	16
2	2	4	4	4
35	28	153	219	122

لاحظ أنك تحتاج في حساب r إلى المقادير التالية :

$$\sum xy , \sum y^2 , \sum x^2 , \bar{y} , \bar{x} , n$$

وقد حسبت في الجدول السابق حيث :

$$\bar{y} = \frac{28}{7} = 7, \bar{x} = \frac{35}{7} = 5$$

وبقية المقادير محسوبة في الجدول.

بتعويض هذه المقادير المحسوبة في معادلة r نجد :

$$r = \frac{153 - 7 \times 5 \times 7}{\sqrt{219 - 7 \times (5)^2}\sqrt{122 - 7 \times 4^2}}$$

$$= \frac{13}{\sqrt{44}\sqrt{10}} = 0.62$$

مثال(4) : الجدول (6) يمثل العلامات النهائية لثمانية طلاب في مقرري الإحصاء (X) والرياضيات (Y).

الجدول (6)

95	85	65	80	45	60	65	85	x
87	82	57	72	52	62	67	77	y

احسب معامل ارتباط بيرسون الخطي بين X, Y.

الحل : لحساب معامل الارتباط r رتب الحل كما في الجدول (7) التالي :

الجدول (7)

x	y	x^2	y^2	x y
85	77	7225	5929	6545
65	67	4225	4489	4355
60	62	3600	3844	3720
45	52	2025	2704	2340
80	72	6400	5148	5760
65	57	4225	3249	3705
85	82	7225	6724	6970
95	87	9025	7569	8265
580	556	43950	39656	41660

ومن الجدول (7) نجد :

$$\bar{x} = \frac{580}{8} = 72.5$$

$$\bar{y} = \frac{556}{8} = 69.5$$

وباستعمال المعادلة :

$$r = \frac{\sum xy - n\bar{x}\bar{y}}{\sqrt{\sum x^2 - n\bar{x}^2}\sqrt{\sum y^2 - n\bar{y}^2}}$$

وبتعويض المقادير المحسوبة في الجدول (7) في معادلة r نجد :

$$r = \frac{41660 - 8 \times 72.5 \times 69.5}{\sqrt{43950 - 8(72.5)^2}\sqrt{39656 - 8(69.5)^2}}$$

$$= \frac{1350}{\sqrt{1900}\sqrt{1014}}$$

$$= \frac{1350}{43.59 \times 31.84} = 0.97$$

مثال (5) :

يمثل الجدول (8) الطول x والوزن y لخمسة طلاب في احدى الكليات :

y كغم	x سم
70	172
71	170
74	165
67	169
58	164

الجدول (8)

احسب معامل ارتباط بيرسون الخطي بين Y, X.

الحل : نرتب خطوات الحل كما في الجدول (9) حيث نحتاج $\sum xy, \sum y^2, \sum x^2, \bar{y}, \bar{x}, n$.

الجدول (9)

x	y	xy	x^2	y^2
172	70	12040	29584	4900
170	71	12070	28900	5041
165	74	12210	27225	5476
169	67	11323	28561	4489
164	58	9512	26896	3364
840	340	57155	141166	23270

$$\bar{x} = \frac{\sum x}{n} = \frac{840}{5} = 168$$

$$\bar{y} = \frac{\sum y}{n} = \frac{340}{5} = 68$$

وبتعويض المقادير المحسوبة في الجدول (9) في معادلة r نجد :

$$r = \frac{57155 - 5 \times 168 \times 68}{\sqrt{141166 - 5(168)^2}\sqrt{23270 - 5(68)^2}}$$

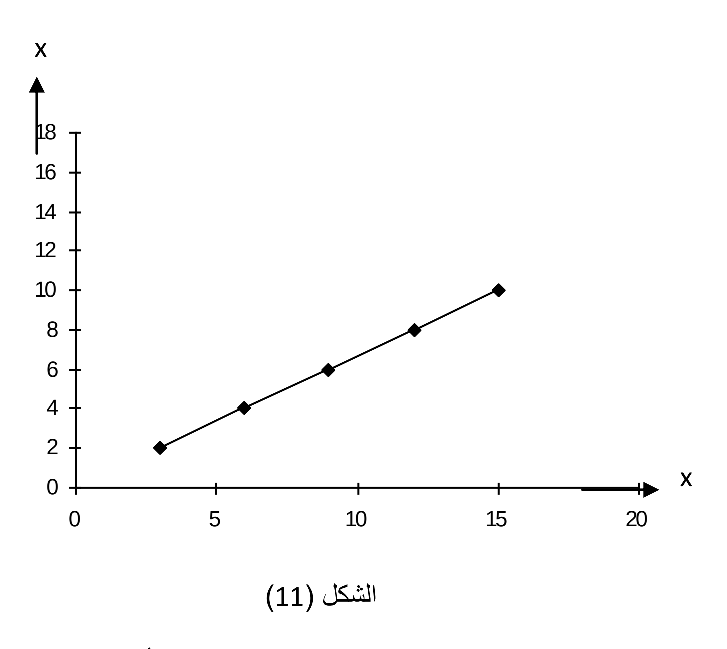
$$= \frac{35}{\sqrt{46}\sqrt{150}}$$

$$= \frac{35}{6.78 \times 12.25} = 0.42$$

10 : 4 تفسير معامل الارتباط

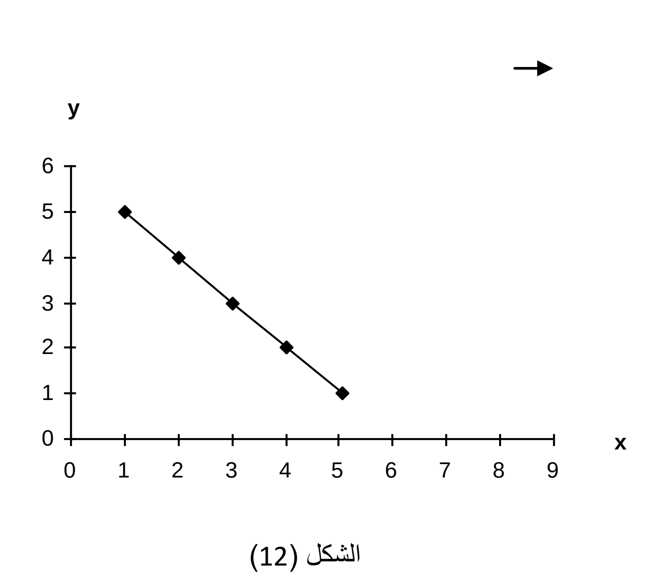
Interpretation of the Correlation Coefficient

1- من الممكن رياضيا اثبات أن قيمة r تتراوح ما بين 1- ,1. فالقيمة 1 = r تعني وجود ارتباط خطي كامل موجب بين المتغيرين،فإذا ازداد أحدهما ازداد الآخر، وإذا كان المتغيران هما Y, X فإن = r 1 تعني أنه إذا رسمنا لوحة الانتشار لقيم X وقيم Y المقابلة لها، لوجدنا أن النقط التي نرصدها في لوحة الانتشار تقع على خط مستقيم ميله موجب، أي أن الزيادة في X يتبعها زيادة في Y.



الشكل (11)

الشكل (11) يعطي مثالا على الارتباط الخطي الكامل الموجب، حيث أنه كلما ازدادت X ازدادت Y، وهذا يعني أن النقط في لوحة الانتشار تقع على خط مستقيم. أما حالة 1- = r فتعني وجود ارتباط خطي كامل سالب بين المتغيرين. وهذا يعني أن النقط في لوحة الانتشار تقع على خط مستقيم ميله سالب أي كلما ازدادت X نقصت Y كما يظهر في الشكل (12).



الشكل (12)

أما r = 0 فتعني عدم وجود ارتباط خطي بين المتغيرين. وفي كثير من الأحيان يكون هناك علاقة خطية بين المتغيرين ، لكنها غير كاملة، وبالتالي يكون r واقعاً ما بين 1-,1 وأن هذه الحالات تدل على وجود ارتباط خطي بين المتغيرين. فإذا كان r قريبا من 1 فهذا يعني وجود ارتباط خطي موجب وبالتالي العلاقة بين المتغيرين طردية وإذا كان r قريبا من 1- فهذا يعني وجود ارتباط خطي سالب وبالتالي تكون العلاقة بين المتغيرين عكسية. والأمثلة (3) ، (4) ، (5) السابقة أمثلة على الارتباط الخطي (لاحظ أنه ارتباط غير كامل).

إذا وقعت جميع نقط لوحة الانتشار على خط مستقيم فقيمة r تساوي 1 إذا كان ميل الخط المستقيم الذي تقع عليه النقط موجبا، أما إذا كان ميل الخط المستقيم الذي تقع عليه النقط سالبا فإن r = -1.

3- إذا لم توجد علاقة خطية بين المتغيرين، يكون r = 0 والعكس صحيح ، إذا كانت r = 0 فإنه لا يوجد علاقة خطية بين المتغيرين.

4- إن وجود ارتباط خطي كامل أو غير كامل بين متغيرين لا يعني السببية، أي أنه إذا كان معامل الارتباط r = 1 بين المتغيرين X,Y فهذا لا يعني أن X سبب في حدوث Y ولكن يعني أن هناك علاقة خطية بين هذين المتغيرين، وقد يكون المتغيران قد تأثرا بمتغير ثالث كان هو السبب في حدوثهما معا، فمثلا : إذا حسبنا معامل الارتباط بين ما تنفقه العائلة كمصروف شهري وما يدخل في خزينة الدولة من رسوم تسجيل الأراضي خلال السنوات 1986 وحتى 1999 وجدنا أن هناك ارتباطا بينهما، وربما يكون معامل الارتباط الخطي بينهما كبيرا. هل نقول إن أحد المتغيرين سبب في الآخر ؟ كلا لكن يمكن أن يعزى الأمر (تأثر المتغيرين الانفاق الشهري، زيادة الدخل من الرسوم) إلى سبب ثالث وعامل ثالث هو زيادة الدخل العام للأسرة وزيادة حجم التحويلات من المغتربين مما أدى إلى زيادة في الإنفاق ونشاط في شراء الأراضي.

5- في كثير من الأحيان لا تكون العلاقة بين المتغيرين خطية. وقد ذكرنا أمثلة على ذلك، إذ من الممكن أن يزداد أحد المتغيرين Y مع ازدياد المتغير الآخر X إلى درجة معينة أو نقطة محددة ثم يتغير الحال ويبدأ Y بالنقصان مع ازدياد X.

في مثل هذه الحالات، لا يصلح استعمال معامل ارتباط بيرسون الخطي ، لأن هذا المعامل يقيس قوة الارتباط الخطي ولا يقيس قوة الارتباط غير الخطي، ومن الممكن معرفة وجود ارتباط خطي بشكل مبدئي بين متغيرين أو عدم وجوده بواسطة لوحة الانتشار كما شرحنا سابقاً.

10 : 5 معامـل الارتبــاط للرتــب (Coefficient of Rank Correlation)

في كثير من الأحيان يصعب قياس متغير ما رقميا ولكن يسهل تعيين رتب للصفة أو المميز المراد دراسته عن هذا المتغير. فمثلا، إذا كان لدينا خمسة أنواع من الشاي وأردنا التمييز بين هذه الأنواع من حيث المذاق، نرى أنه يسهل على الذواقة ترتيب أنواع الشاي من الدرجة الأولى في المذاق حتى الدرجة الخامسة، ولكن ربما يصعب عليه إعطاء قيم عددية لكل نوع من أنواع الشاي السابقة من حيث المذاق. ينطبق هذا التحليل على كثير من المسائل في الاقتصاد والإدارة والتربية وغيرها.

من أهم معاملات الارتباط للرتب معامل سبيرمان Spearman's Coefficient of rank correlation

الذي يعرف بالقانون :

$$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

حيث n تساوي عدد أزواج البيانات (X, Y) و d تساوي الفرق بين رتب Xو Y.

نلاحظ من هذا التعريف أنه يمكن حساب قيمة r_s إذا عرفت الرتب أو إذا عرفت البيانات التي يمكن ترتيبها. ويصلح هذا المعامل بوجه خاص إذا كان عدد أزواج البيانات ما بين 25 و 30 أو أقل.

مثال (6) :

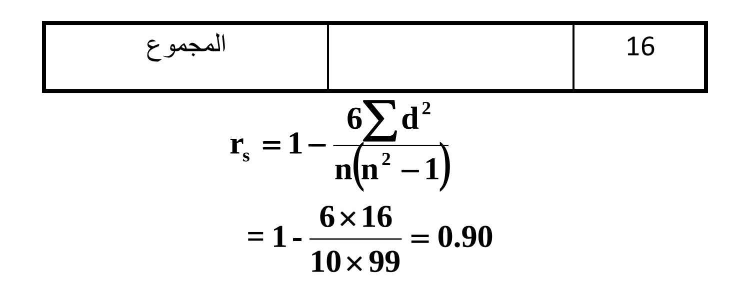
احسب معامل سبيرمان للارتباط بالرتب بين المعدلات التالية لعشرة طلاب في شهادة الدراسة الثانوية والفصل الجامعي الأول :

معدل الطالب في شهادة الدراسة الثانوية X	77.5	85.5	79.9	88.8	93.5	85.0	94.1	90.5	87.2	89.8
معدل الطالب في نهاية الفصل الجامعي الأولY	61.0	72.8	65.5	71.0	80.0	74.5	82	81	76.1	78.0

نرتب المعدلات X بحيث نعطي الرتبة 1 لأعلى معدل وهكذا، ونرتب المعدلات Y بالمثل ثم نجد الفرق بين رتبتي كل طالب كما هو موضح في الجدول التالي الذي نحسب منه معامل سبيرمان r_s :

رتب X	رتب Y	الفرق بين الرتب d	d^2
10	10	0	0
7	7	0	0
9	9	0	0
5	8	-3	9
2	3	-1	1
8	6	2	4
1	1	0	0
3	2	1	1
6	5	1	1
4	4	0	0

المجموع		16



$$r_s = 1 - \frac{6\sum d^2}{n\left(n^2 - 1\right)}$$

$$= 1 - \frac{6 \times 16}{10 \times 99} = 0.90$$

في المثال (6) لم تظهر معدلات متساوية في أي من البيانات X أو Y، أما إذا ظهرت بيانات متساوية (Tied data) فيكون تعيين الرتب لهذه البيانات باتباع الخطوتين :

1- نرتب البيانات كما لو أن ليس فيها بيانات متساوية.

2- نأخذ الوسط الحسابي لرتب كل مجموعة من البيانات المتساوية ونعتبر هذا الوسط الحسابي رتبة كل بيان في هذه المجموعة.

مثال (7) : عين الرتب للعلامات التالية :

70, 79, 63, 70, 63, 65, 63, 57, 53, 57, 45.

الحل : نكتب العلامات تنازليا ونعين الرتب كما يلي :

45,	53,	57,	57,	63,	63,	63,	65,	70,	70,	79
(11)	(10)	(9)	(8)	(7)	(6)	(5)	(4)	(3)	(2)	(1)

نجد الوسط الحسابي لرتب العلامات 70 و 70 وهو 2.5.

ونجد الوسط الحسابي لرتب العلامات 63, 63, 63 وهو 6، ونجد الوسط الحسابي لرتب العلامات 57, 57 وهو 8.5 وتكون الرتب كما في الجدول (10).

جدول (10)

70	79	63	70	63	65	63	57	53	57	45	العلامة
2.5	1	6	2.5	6	4	6	8.5	10	8.5	11	الرتبة

10 : 6 دلالة معامل الارتباط

Significance of Correlation Coefficient

إذا كان r قريبا من ± 1 كان هناك علاقة خطية قوية بين المتغيرين وإذا كان $r = 0$ لم يكن هناك علاقة خطية ، أما قيم r المتوسطة الأخرى فيجب اختبار دلالة معامل الارتباط لها.

ولإجراء هذا الاختبار نعتبر المجتمع ذا البعدين X, Y الذي أخذت منه المجموعة المكونة من n من الأزواج المرتبة ونفترض أن معامل الارتباط لهذا المجتمع ρ فعندها يكون r تقديرا للمعامل ρ.

الآن ، لو أخذنا جميع العينات الممكنة ذات الحجم n والمكونة من أزواج المشاهدات وافترضنا أن المجتمع الأصلي يخضع للتوزيع الطبيعي ثم حسبنا توزيع المعاينة لمعامل الارتباط r فلا نحصل على توزيع من التوزيعات المألوفة لدينا ولكن إذا افترضنا $\rho = 0$ فإننا نحصل على اقتران لـ (r) يخضع لتوزيع معلوم كما يأتي :

نظرية (1) : إذا أخذت جميع العينات الممكنة ذات الحجم n من مجتمع ذي بعدين خاضع للتوزيع الطبيعي ومعامل ارتباطه $\rho = 0$ ، وإذا كان r يعبر عن معاملات ارتباطات تلك العينات فإن :

$$\mathbf{t = \frac{r}{\sqrt{(1 - r^2)/(n - 2)}}}$$

يخضع لتوزيع t على (n-2) درجات حرية. والآن ، باستعمال هذه النظرية، يمكننا تحديد فيما إذا كان معامل الارتباط r لعينة حجمها n دالا على وجود علاقة خطية بين المتغيرين. يكون هذا التحديد باختبار الفرضية $\rho = 0$ على مستوى دلالة معين.

مثال (8) : في المثال (4) هل تدل قيمة r على وجود ارتباط ذي دلالة بين علامة الطالب في الإحصاء وعلامته في الرياضيات على مستوى دلالة 5%؟

1- نختبر الفرضية $\mathbf{H_0 : \rho = 0}$

2- مقابل الفرضية $\mathbf{H_1 : \rho \neq 0}$

3- $\alpha = \mathbf{0.05}$.

4- نحسب الإحصاء (إحصاء الاختبار والحسابات).

$$\mathbf{t = \frac{r}{\sqrt{(1 - r^2)/(n - 2)}} = \frac{0.97}{\sqrt{(1 - 0.941)/(8 - 2)}}}$$
$$\mathbf{= \frac{0.97}{0.099} = 9.8}$$

5- منطقة الرفض نجد من جدول t :

$$\mathbf{t[0.975;6] = 2.447}$$
$$\mathbf{t[0.025;6] = -2.447}$$

أرفض H_0 إذا كان $\mathbf{|t| > 2.45}$

6- المقارنة : بما أن t المحسوبة أكبر من $\mathbf{t\left[1 - \frac{\alpha}{2}; n - 2\right]}$ أي $9.8 > 2.45$

نرفض H_0 ونعتبر وجود ارتباط ذي دلالة بين المتغيرين.

مثال (9) : هل تدل قيمة r في مثال (5) على وجود ارتباط ذي دلالة بين الطول X والوزن Y على مستوى دلالة 1%.

الحل :

1- نختبر الفرضية $\mathbf{H_0 : \rho = 0}$

2- مقابل الفرضية $\mathbf{H : \rho \neq 0}$

3- مستوى الدلالة $\boldsymbol{\alpha = 0.01}$.

4- إحصاء الاختبار هو :

$$\mathbf{t = \frac{r}{\sqrt{\left(1-r^2\right)/\left(n-2\right)}}}$$

5- منطقة الرفض : من جدول t نجد :

$$\mathbf{t[0.995;3] = 5.841}$$
$$\mathbf{t[0.005;3] = -5.841}$$

أرفض H_0 إذا كان $\mathbf{|t| > 5.841}$

6- الحسابات :

$$\mathbf{t = \frac{0.42}{\sqrt{\left(1-(0.42)^2\right)/5-2}}}$$
$$\mathbf{= \frac{0.42}{\sqrt{0.274}} = \frac{0.42}{0.523} = 0.8}$$

7- المقارنة : بما أن قيمة t المحسوبة 0.8 لا تقع في منطقة الرفض فلا يوجد دليل على رفض $\mathbf{H_0 : \rho = 0}$ والاستنتاج أن ليس هناك ارتباط بين المتغيرين.

بما أن نظرية (1) تتحقق فقط في حالة $\rho = 0$ فإنه لا يمكن استعمالها لاختبار الفرضيات مثل $H_0 : \rho = \rho_0$ أو إيجاد فترات ثقة ل (ρ)، ولتحقيق هذا الهدف عرف فيشر Fisher الإحصاء (Z فيشر) الذي يحقق النظرية التالية :

نظرية (2) :

إذا أخذت جميع العينات ذات الحجم n من مجتمع ذي بعدين معامل الارتباط بينهما ρ، وإذا عرفنا الإحصاء :

$$Z = \frac{1}{2}\ln\frac{1+r}{1-r}$$

فإن Z يقرب من التوزيع الطبيعي ذي المعدل μ_Z والانحراف المعياري σ_Z حيث:

$$\mu_Z = \frac{1}{2}\ln\frac{1+\rho}{1-\rho}\ ,\ \sigma_Z = \frac{1}{\sqrt{n-3}}$$

لاحظ أن (Z فيشر) ليست هي Z التي استعملناها رمزا للطبيعي المعياري.

نستعمل نظرية (2) في إيجاد فترة ثقة 100%(1-α) لمعامل الارتباط ρ وذلك بإيجاد فترة ثقة لـ μ_Z ثم تحويلها إلى فترة ثقة لـ (ρ) كما يلي :

$$P\left(-z_{1-\frac{\alpha}{2}} \le \frac{Z-\mu_Z}{\sigma_Z} \le z_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

ومنه نستنتج أن :

$$\left(Z - z_{1-\frac{\alpha}{2}}\sigma_Z , Z + z_{1-\frac{\alpha}{2}}\sigma_Z\right)$$

هي فترة ثقة 100%(1-α) للمعلمة μ_Z وبقراءة ما يقابل هذه الفترة من جدول تحويل r إلى Z فيشر نجد الفترة المطلوبة للمعلمة ρ.

مثال (10) :

أوجد فترة ثقة 95% لمعامل الارتباط ρ إذا أعطت عينة عشوائية حجمها n =12 من الأزواج المرتبة (X, Y) معامل ارتباط r = 0.7.

الحل : من جدول التوزيع الطبيعي المعياري $z_{0.975} = 1.96$ ومن النظرية (2) :

$$\sigma_Z = \frac{1}{\sqrt{n-3}} = 0.33$$

من جدول Z فيشر نجد : $r = 0.7$ يقابلها $Z = 0.867$

إذا فترة الثقة للمعلمة μ_Z هي :

$(0.867 - 1.96 \times 0.33 , 0.867 + 1.96 \times 0.33)$

أي $(0.22 , 1.51)$

والآن : $Z = 0.22$ يقابلها في جدول Z فيشر $r = 0.22$ ، $Z = 1.51$ يقابلها $r = 0.907$.

إذاً فترة ثقة 95% لمعامل الارتباط ρ هي $(0.22, 0.907)$.

وتستعمل النظرية (2) في اختبار الفرضيات مثل $H_0 : \rho = \rho_0$ حيث $\rho_0 \neq 0$ وذلك بإتباع الخطوات المألوفة لاختبار الفرضيات التي يخضع توزيع إحصاء الاختبار فيها للتوزيع الطبيعي.

مثال (11) :
في مثال (10) اختبر $H_0 : \rho = 0.6$ مقابل $H_1 : \rho \neq 0.6$ على مستوى دلالة $\alpha = 0.05$.

الحل :

1- نختبر الفرضية $H_0 : \rho = 0.6$

2- مقابل الفرضية $H_1 : \rho \neq 0$

3- مستوى الدلالة $\alpha = 0.05$.

4- إحصاء الاختبار (حسب نظرية (2)). هو Z فيشر الذي يخضع لتوزيع طبيعي معدله μ_Z وانحرافه المعياري σ_Z .

5- منطقة الرفض : ارفض H_0 إذا كانت القيمة المعيارية للإحصاء Z فيشر أصغر من $-z_{0.975} = -1.96$ أو أكبر من $z_{0.975} = 1.96$.

6- الحسابات : $r = 0.7$ يقابلها من جدول Z فيشر $Z = 0.867$.

$\rho = 0.6$ يقابلها $\mu_Z = 0.693$ ، $\sigma_Z = \frac{1}{\sqrt{12-3}} = 0.33$

إذاً القيمة المعيارية لـ Z فيشر هي :

$$Z = \frac{0.867 - 0.693}{0.333} = 0.522$$

(7) المقارنة :

بما أن $0.522 \not> 1.966$

إذاً لا نرفض الفرضية الصفرية $H_0 : \rho = 0.6$.

مثال (12) :

في مثال (11) هل نرفض $H_0 : \rho = 0.2$ مقابل $H_1 : \rho \neq 0.2$.

الحل : يبقى الحل نفسه كالسابق ولكن الآن $\rho = 0.2$ يقابلها $\mu_Z = 0.203$ إذا القيمة المعيارية لـ Z هي :

$$Z = \frac{0.867 - 0.203}{0.333} = 1.99$$

وبما أن 1.96 < 1.99 ، إذاً نرفض $H_0 : \rho = 0.2$ لصالح $\rho > 0.2$.

ملاحظة : نلاحظ أن بالامكان حساب Z من معادلتها بدلالة r أو استعمال جدول تحويل r إلى Z ، ويماثل ذلك العلاقة بين ρ و μ_Z كما جاء في منطوق نظرية (2).

الانحدار Regression

10 : 7 مفهوم الانحدار Concept of Regression

من أهم التطبيقات الإحصائية في الاقتصاد والإدارة والعلوم البحته والعلوم التربوية وغيرها مسألة التنبؤ، وهو يبنى على وجود علاقة بين متغيرين، كأن تعرف العلاقة التي تربط المتغيرين Y , X مثلا وتريد أن تقدر أو تتنبأ قيمة Y التي تقابل قيمة معينة للمتغير X.

لقد سبق أن درست مفهوم الارتباط والارتباط الخطي ، وعرفت كيف تحسب قيمة معامل الارتباط الخطي بين متغيرين. بذلك تعرفت على طريقة تقيس بها قوة العلاقة الخطية بين متغيرين، هل هذه العلاقة طردية أم عكسية. وقد درست لوحة الانتشار التي تم فيها توضيح هذه المفاهيم.

وفي هذا البند نحاول الإجابة عن السؤال التالي :

إذا كانت هناك علاقة خطية بين متغيرين فكيف نعبر عن هذه العلاقة بواسطة معادلة تربط بين المتغيرين ؟

تسمى المعادلة الخطية التي تعطيك أحد المتغيرين بدلالة المتغير الآخر معادلة الانحدار الخطي. والتنبؤ واحد من أهم أغراض دراسة الانحدار، وهو تنبؤ قيمة متغير ما إذا عرفت قيمة متغير آخر أو متغيرات أخرى.

سنقصر الشرح على مفهوم تنبؤ قيمة متغير واحد إذا عرفت قيمة متغير آخر. فعلى سبيل المثال ، افرض أنك تريد تنبؤ (تقدير) معدل الطالب في السنة الجامعية الأولى بناء على معرفتك معدله في امتحان شهادة الدراسة الثانوية، قبل أن يلتحق الطالب بالدراسة الجامعية.

إن تنبؤ معدل الطالب في السنة الجامعية الأولى لا يعتمد فقط على معدل ذلك الطالب في امتحان شهادة الدراسة الثانوية، بل على عوامل اجتماعية واقتصادية وحياتية أخرى، ولكن لتبسيط شرح المثال نقصر الأمر على متغيرين اثنين كالآتي :

إذا فرضت أن معدل أحد الطلبة في شهادة الدراسة الثانوية هو X ومن معرفتنا لقيمة X نريد أن نتنبأ معدل الطالب في السنة الجامعية الأولى Y. إذا علمنا أن هناك علاقة خطية بين Y, X وأمكن التعبير عن هذه العلاقة بمعادلة خطية تربط بين Y, X فتعطي قيمة Y بدلالة X فإن هذه المعادلة الخطية تسمى خط انحدار Y على X، ويسمى X المتغير المستقل ويسمى Y المتغير التابع. إذا عرفت قيمة معينة للمتغير X نعوّض هذه القيمة في معادلة الانحدار فنجد تنبؤا $\hat{Y}$ للمتغيرY المقابلة للقيمة المعطاة. هناك أمثلة عديدة تبين الحاجة إلى تنبؤ قيمة متغير من معرفة قيمة متغير آخر، مثل تنبؤ كميات المبيعات من سلعة إذا عرفت تكاليف الدعاية عن تلك السلعة.

10 : 8 معادلة انحدار متغير على متغير آخر

Regression Equation of One Variable on another

إذا كان لدينا عينة من الأزواج المرتبة :

$$(x_1, y_1).(x_2, y_2),......(x_n, y_n)$$

حيث $x_1, x_2, ...x_n$ قيم لمتغير X , وحيث $y_1, y_2, ...y_n$ القيم المقابلة لها للمتغير Y ، ورصدنا هذه النقاط على المستوى xy نحصل على لوحة الانتشار، ونتمكن من الحكم فيما إذا كان من المعقول تطبيق خط مستقيم على شكل الانتشار أم لا.

فإذا فرضنا أن هناك علاقة خطية بين المتغيرين X، Y أمكن التعبير عن هذه العلاقة بالمعادلة:

$$Y = A + BX + e$$

ويقابل هذه المعادلة للمشاهدة التي رقمها i المعادلة:

$$Y_i = A + BX_i + e_i$$

حيث e_i هي الأخطاء $i = 1, ..., n$ وهي متغيرات عشوائية مستقلة ويخضع كل منها للتوزيع الطبيعي $N(0, \sigma^2)$ وهذه الأخطاء غير مشاهدة.

والمطلوب هو تقدير A، B لكي نستطيع تقدير Y_i المقابلة للمتغير X_i.

نفرض أن a هو تقدير A، b هو تقدير B ولذلك يكون:

$$\hat{Y} = a + bx$$

هو خط انحدار Y على X الذي حصلنا عليه باستعمال a، b ويكون $\hat{Y}$ هي القيمة التقديرية (المتنبأة) ويكون $e = Y - \hat{Y}$ هو الخطأ في التقدير.

من المنطقي أن نبحث عن أحسن خط مستقيم يمكن أن يطبق على لوحة الانتشار.

إن إحدى الطرق التي تستعمل لإيجاد ذلك الخط المستقيم هي طريقة المربعات الصغرى، وتتلخص هذه الطريقة في إيجاد قيم b, a التي تجعل مجموع مربعات الأخطاء أصغر ما يمكن.

تعريف (2) :
طريقة المربعات الصغرى هي طريقة تطبيق خط مستقيم على مجموعة من النقاط $(x_1, y_1).(x_2, y_2),......(x_n, y_n)$ بحيث يكون مجموع مربعات الأخطاء، أي $\sum(y - \hat{y})^2$ أصغر ما يمكن، وذلك بحل المعادلتين التا ليتين :

$$an + b\sum x = \sum y$$
$$a\sum x + b\sum x^2 = \sum xy$$

وتسمى هاتان المعادلتين القانونيتين وبحل هاتين المعادلتين نجد :

$$b = \frac{\sum xy - n\bar{x}\,\bar{y}}{\sum x^2 - n\bar{x}^2}$$
$$a = \bar{y} - b\bar{x}$$

حيث $\bar{x}$ هو الوسط الحسابي للمشاهدات $\bar{y}, x_1, x_2, ...x_n$ هو الوسط الحسابي للمشاهدات $y_1, y_2, ..., y_n$ وبتعويض قيمتي b,a في المعادلة $\hat{Y} = a + bx$ نحصل على معادلة انحدار Y على X.

مثال (13) :
بالإشارة إلى مثال (4) أوجد معادلة انحدار علامة الرياضيات (Y) على علامة الإحصاء (X). ثم أوجد تقديرا لعلامة الرياضيات لطالب حصل في الإحصاء على 80 ثم أوجد الخطأ في التقدير.

الحل : افرض معادلة انحدار Y على X بطريقة المربعات الصغرى هي :

$$\hat{y} = a + bx$$

احسب قيم b , a من المعادلتين ، وذلك بتعويض المقادير المحسوبة في الجدول (7).

$$b = \frac{41660 - 8 \times 72.5 \times 69.5}{43950 - 8(72.5)^2}$$
$$= \frac{1350}{1900} = 0.71$$
$$a = \bar{y} - b\bar{x} = 69.5 - 0.71 \times 72.5$$
$$= 18.03$$

إذن تكون معادلة انحدار Y على X هي :

$$\hat{y} = 18.03 + 0.71x$$

ويكون تقدير علامة الرياضيات إذا كانت علامة الإحصاء80 بتعويض x= 80 في المعادلة وإيجاد قيمة $\hat{y}$ وهي :

$$\hat{y} = 18.03 + 0.71 \times 80$$
$$= 74.8 \approx 75$$

وهو تنبؤ علامة الرياضيات.

أما الخطأ في التقدير فهو :

$$e = y - \hat{y} = 72 - 75 = -3$$

حيث 72 هي القيمة المشاهدة المقابلة للعلامة 80، والعلامة 75 هي القيمة المتنبأة.

مثال (14) :

في مثال (5) أوجد معادلة انحدار الوزن Y على الطول X.

الحل : من حل المثال في جدول (8) نجد :

$$b = \frac{\sum xy - n\bar{x}\bar{y}}{\sum x^2 - n\bar{x}^2}$$
$$= \frac{57155 - 5 \times 168 \times 68}{141166 - 5 \times (168)^2}$$
$$= \frac{35}{46} = 0.76$$

$$a = \bar{y} - b\bar{x}$$
$$= 68 - 0.76 \times 168 = -59.8$$

إذا معادلة انحدار الوزن Y على الطول X هي :

$$\hat{y} = -59.8 + 0.76x$$

إذا كان طول طالب 180سم فإننا نقدر وزنه :

$$\hat{y} = -59.8 + 0.76 \times 180$$
$$= 77$$

مثال (15) :

(أ) أوجد معادلة انحدار Y على X للبيانات في الجدول التالي :

5	8	12	9	10	4	x
4	5	11	8	6	2	y

(ب) بماذا تقدر قيمة y المقابلة للقيمة x = 9 (جـ) ما الخطأ في تقدير y عندما تكون x = 9.

الحل :

(أ) رتب الحل كما في الجدول (11).

الجدول (11)

	X	Y	XY	X^2
	4	2	8	16
	10	6	60	100
	9	8	72	81
	12	11	132	144
	8	5	40	64
	4	4	20	25
المجموع	48	36	332	430

معادلة انحدار Y على X هي :

$$\hat{y} = a + bx$$

حيث $b = \dfrac{\sum xy - n\bar{x}\bar{y}}{\sum x^2 - n\bar{x}^2}$

$$a = \bar{y} - b\bar{x}$$

عوض القيم المطلوبة في حساب a, b من الجدول (11) فتجد :

$$\bar{x} = \frac{48}{6} = 8 \ , \ \bar{y} = \frac{36}{6} = 6$$

$$b = \frac{332 - 6 \times 8 \times 6}{430 - 6(8)^2} = 0.96$$

$$a = 6 - 0.96 \times 8 = -1.68$$

إذن ، معادلة انحدار Y على X هي :

$$\hat{y} = 1.68 + 0.96x$$

ب) القيمة التقديرية (المتنبأة) للمتغيرY عندما x = 9 هي :

$$\hat{y} = 1.68 + 0.96 \times 9 = 6.96$$

وذلك بتعويض x = 9 في معادلة الانحدار.

جـ) من الجدول المعطى في المثال والذي يبين المشاهداتX والمشاهدات المقابلة لها Y نجد أن المشاهدة التي تقابل x = 9 هي y = 8 . وكما وجدنا في (ب)

تقدر قيمة Y بالمقدار $\hat{y} = 6.96$.

إذا الخطأ في التقدير هو $e = y - \hat{y}$

= 8 - 6.96 = 1.04

مثال (16) : يعطي الجدول التالي علامات 12 طالبا في الامتحان الأول X والامتحان الثاني Y.

Y	X
20	18
11	14
14	10
16	15
10	7
10	12
17	13
11	8
12	9
20	17
18	15
12	12

1- ارسم لوحة الانتشار.

2- أوجد معادلة خط انحدار Y على X.

3- أخذ طالب في الاختبار الأول العلامة 16 وغاب عن الاختبار الثاني، ما هي العلامة التقديرية التي يحصل عليها الطالب في الامتحان الثاني ؟

4- ما هو الخطأ في تقديرك للعلامة في الاختبار الثاني إذا حصل الطالب على 10 في الاختبار الأول موضحا ذلك بالرسم.

الحل :

1- نحصل على لوحة الانتشار برصد النقاط (x, y) المعطاة في الجدول كما يظهر في الشكل (13).

2- نجري الحسابات كما في الجدول (12).

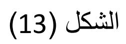
الجدول (12)

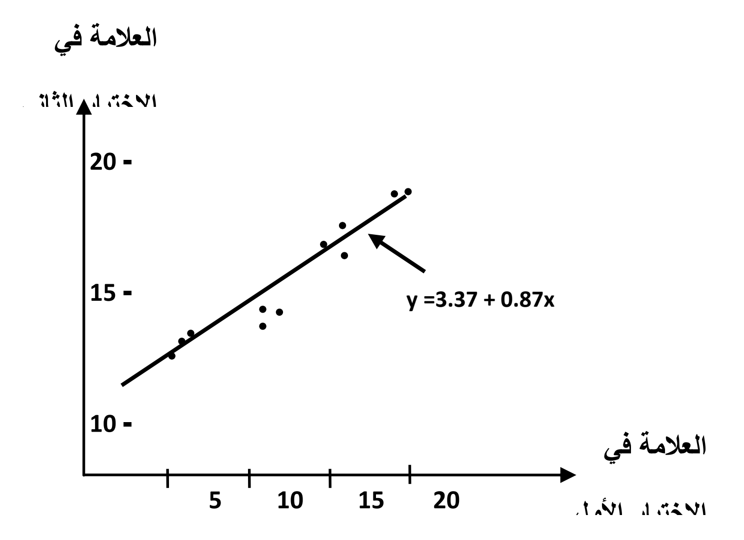
	X	Y	XY	X^2
	18	20	360	324
	14	11	154	196
	10	14	140	100
	15	16	240	225
	7	10	70	49
	12	10	120	144
	13	17	221	169
	8	11	88	64
	9	12	108	81
	17	20	340	289
	15	18	270	225
	12	12	144	144
المجموع	150	171	2255	2010

ومنه نجد :

$$\bar{x} = \frac{150}{12} = 12.5$$

$$\bar{y} = \frac{171}{12} = 14.25$$


الشكل (13)



وبالتعويض في المعادلة :

$$b = \frac{\sum xy - n\bar{x}\bar{y}}{\sum x^2 - n\bar{x}^2}$$

نجد :

$$b = \frac{2255 - 12 \times 12.5 \times 14.25}{2010 - 12(12.5)^2}$$

$$b = \frac{2255 - 2137.5}{2010 - 1875} = \frac{117.5}{135} = 0.87$$

وبالتعويض في المعادلة :

$$a = \bar{y} - b\bar{x}$$

نجد : $a = \frac{171}{12} - 0.87 \times 12.5 = 14.25 - 10.875$

تقريباً $= 3.37$

وبالتعويض في معادلة خط الانحدار $\hat{y} = a + bx$

نجد $\hat{y} = 3.37 + 0.87x$ هي معادلة خط انحدار Y على X.

3- نعوض قيمة x = 16 في معادلة خط الانحدار فنجد العلامة التقديرية في الامتحان الثاني :

$$\hat{y} = 3.37 + 0.87 \times 16$$

تقريبا 17 = 17.29 =

4- نعوض x = 10 في معادلة الانحدار فنجد

$$\hat{y} = 3.37 + 0.87 \times 10 = 1207 \approx 12$$

ويكون الخطأ في التقدير $\hat{e} = 14 - 12 = 2$.

10 : 9 تقدير التباين لخط ا لانحدار

Estimate of the Variance of the Regression Line

إن الانحرافات المفردة للمشاهدات y_i عن قيم خط الانحدار $\hat{y}_i = a + bx_i$ تسمى البواقي (أو الأخطاء) ويعبر عنها بالرمز $\hat{e}_i$ ، حيث $\hat{e}_i = y_i - a - bx_i = y_i - \hat{y}_i$ حيث i = 1, 2, ... , n .

ومع أن بعض قيم $\hat{e}_i$ موجبة وبعضها سالب إلا أن مجموعها دائما صفر. أما مجموع مربعات البواقي فهي لا تساوي صفرا وهي ذات أهمية في تقدير σ^2 تباين توزيعات الأخطاء e_i التي افترضناها سابقا. وبعض الأحيان نعبر عن مجموع مربعات البواقي بمجموع المربعات للخطأ ويعطى بالقاعدة :

قاعدة (1) مجموع مربعات البواقي Residual Sum of Squares

أو مجموع المربعات للخطأ Sum of Squares due to Error هو

$$SSE=\sum_{i=1}^{n}\hat{e}_i^2$$

$$=S_{yy}-\frac{S_{Xy}^2}{S_{xx}}$$

$$S_{yy}=\sum(y-\bar{y})^2=\sum y^2-n\bar{y}^2$$

$$S_{xx}=\sum(x-\bar{x})^2=\sum x^2-n\bar{x}^2$$

$$S_{xy}=\sum(x-\bar{x})(y-\bar{y})=\sum xy-n\overline{xy}$$

أما تقدير σ^2 أي تقدير تباين خط انحدار Y على X فيعطى بالمعادلة :

$$S_e^2=\frac{SSE}{n-2}$$

أما الجذر التربيعي الموجب لهذا التقدير فيسمى "تقدير الخطأ المعياري".

$$S_e=\sqrt{\frac{SSE}{n-2}}$$

مثال (17) :

أوجد معادلة خط انحدار Y على X ثم أوجد الخطأ المعياري للتقدير للبيانات:

x	1	4	3	7	5
y	3	5	4	10	8

الحل : (أ) معادلة الانحدار هي : $\hat{y}=a+bx$

حيث $b=\frac{\sum xy-n\bar{x}\,\bar{y}}{\sum x^2-n\bar{x}^2}$

$$a=\bar{y}-b\bar{x}$$

ولحساب هذه المقادير رتب خطوات الحل كما في الجدول (13).

جدول (13)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
x	y	xy	x^2	$\hat{y}$	$(y-\hat{y})$	$(y-\hat{y})^2$	y^2
1	3	3	1	2.25	0.751	0.5625	9
4	5	20	16	6.00	-1.0016	1.0000	25
3	4	12	9	4.75	-0.759	0.5625	16
7	10	70	49	9.75	0.2549	0.0625	100
5	8	40	25	7.25	0.7525	0.5625	64
20	30	145	100			2.75	214

إذن $\mathbf{n = 5, \bar{y} = \frac{30}{5} = 6, \bar{x} = \frac{20}{5} = 4}$ وبالتعويض من الجدول نجد :

$$\mathbf{b = \frac{145 - 5 \times 4 \times 6}{100 - 5 \times (4)^2} = \frac{145 - 120}{100 - 80} = 1.25}$$

$$a = 6 - 1.25 \times 4 = 1$$

إذن معادلة الانحدار هي $\mathbf{\hat{y} = 1 + 1.25x}$

ولحساب الخطأ المعياري للتقدير تحتاج إلى حساب $\hat{y}$ المقابلة لكل قيمة x في معادلة الانحدار فتجد $\hat{y}$ المقابلة لها ، كما يلي :

$\hat{y}$ المقابلة للقيمة $x = 1$ هي :

$$\mathbf{\hat{y} = 1 + 1.25 \times 1 = 2.25}$$

$\hat{y}$ المقابلة للقيمة $x = 4$ هي :

$$\hat{y} = 1 + 1.25 \times 4 = 6$$

وهكذا تجد القيم $\hat{y}$ المقابلة للقيم x في الجدول، ونضع هذه القيم في العمود (5) في الجدول.

ثم نجد الانحرافات $(y - \hat{y})$ ونضعها في العمود (6) ونربع الانحرافات ونضعها في العمود (7)، نجمع مفردات العمود (7) ونقسم على2 - n أي 3 = 2 - 5 ونجد الجذر التربيعي للناتج فنحصل على الخطأ المعياري للتقدير.

$$S_e = \sqrt{\frac{\sum (y - \hat{y})^2}{n - 2}} = \sqrt{\frac{2.75}{5 - 2}} = 0.96$$

لاحظ أننا حسبنا S_e من التعريف ولذلك كان العمل مضنيا حيث اضطررنا لحساب $\hat{y}_i$ لجميع ,i = 1 n , ... ثم وجدنا $\sum (y_i - \hat{y}_i)^2$ ، ولكن إذا استعملنا القاعدة (1) السابقة فإننا نحتاج إلى حساب y_i^2 ونضعها في العمود (8) في الجدول ونحسب :

$$S_{yy} = 214 - 5(6)^2 = 34$$
$$S_{xx} = 100 - 5(4)^2 = 20$$
$$S_{xy} = 145 - 5(6)(4) = 25$$
$$SSE = 34 - \frac{(25)^2}{20} = 2.75$$

ومنه

$$S_e = \sqrt{\frac{SSE}{n - 2}} = \sqrt{\frac{2.75}{3}} = 0.96$$

ويمكن حساب S_e من النظرية التالية :

نظرية (3) :

$$S_e = \sqrt{\frac{\sum y^2 - a\sum y - b\sum xy}{n - 2}}$$

لاحظ في تطبيق هذه النظرية أنك تستعمل المقادير المحسوبة أصلا لحساب معادلة الانحدار. وهذه المقادير هي $a, b, \sum y, \sum xy$ وتحتاج إلى حساب $\sum y^2$ إضافة إلى ما تكون قد قمت بحسابه.
ففي مثالنا السابق يكون الحل كالتالي :

x	y	xy	x^2	y^2
1	3	3	1	9
4	5	20	16	25
3	4	12	9	16
7	10	70	49	100
5	8	40	25	64
20	30	145	100	214

لقد تم حساب b, a سابقا وكانت b = 1.25 ، a = 1.

من النظرية :

$$S_e = \sqrt{\frac{214 - 1 \times 30 - 1.25 \times 145}{5 - 2}}$$
$$= \sqrt{\frac{2.75}{3}} = 0.96$$

لاحظ سهولة الحل حسب قاعدة (1) أو حسب نظرية (3) وذلك لعدم حاجتنا لحساب $\hat{y}_i$.

بعد أن وجدنا S_e، ينشأ لدينا الحاجة إلى تقدير الانحراف المعياري للخطأ في التنبؤ :

(1) إذا علمت $x = x^*$ وكان التنبؤ بالمعدل.

إذا علمت قيمة محددة للمتغير المستقل ولتكن *x فإنك تتنبأ قيمة المتغير التابع لتكون في المعدل مساوية إلى (a + bx*).

ولهذا التنبؤ خطأ معياري يعتمد على الانحراف المعياري للمتغيرY . وما تحتاجه هو تقدير للانحراف المعياري للخطأ في التنبؤ وهو يعطى بالمقدار :

$$S_e\sqrt{\frac{1}{n}+\frac{(x^*-\bar{x})^2}{S_{xx}}}$$

حيث S_e هي تقدير الخطأ المعياري ، و $\bar{x}$ هو الوسط الحسابي للقيم، *x هي القيمة المعطاة والمراد إيجاد التنبؤ $\hat{y}$ المقابل لها ،n هو عدد أزواج المشاهدات المعطاة، $S_{xx}=\sum(x-\bar{x})^2$.

(2) إذا علمت x وكان تنبؤ y بقيمة مفردة أي إذا علمت أن x = x* قيمة محددة للمتغير المستقل وتنبأت المتغير التابع بقيمة واحدة فإن تقدير الانحراف المعياري للخطأ في التنبؤ هو :

$$S_e\sqrt{1+\frac{1}{n}+\frac{(x^*-\bar{x})^2}{S_{xx}}}$$

حيث الرموز كما ذكرنا سابقا في (1).

مثال (18) : في مثال (17) أوجد تقديرا للخطأ المعياري لمعدل التنبؤ إذا كان x = 6 ثم أوجد تقديرا للخطأ المعياري عندما تتنبأ مشاهدة واحدة y والتي تقابل x = 6.

الحل : عندما x = 6 فإن المعدل يكون A+B×6 وهذا يقدر بالقيمة $\hat{y}=a+b\times 6$ وتقدير الخطأ المعياري يكون :

$$S_e\sqrt{\frac{1}{n}+\frac{(x^*-\bar{x})^2}{S_{xx}}}=0.96\sqrt{\frac{1}{5}+\frac{(6-4)^2}{20}}$$
$$=0.96\sqrt{0.4}=0.61$$

أما تقدير الخطأ المعياري عند تنبؤ قيمة مفردة فهو

$$S_e\sqrt{1+\frac{1}{n}+\frac{(x^*-\bar{x})^2}{S_{xx}}}=0.96\sqrt{1+\frac{1}{5}+\frac{(6-4)^2}{20}}$$
$$=0.96\sqrt{1.4}=1.16$$

10 : 10 فترات الثقة واختبار الفرضيات حول A وحول B :

Confidence Intervals and Testing Hypotheses About A and B:

إن معادلة انحدار Y على X تستعمل في تنبؤ قيمة Y المرتبطة بقيمة X المعلومة، وأن التقديرين b , a هما تقديران للمعلمتين B , A على التوالي، ويعتمدان على العينة المكونة من n من أزواج المشاهدات ويخضعان لتغير المعاينة. ولمعرفة مدى دقة تنبؤ قيمة Y باستعمال خط الانحدار عرفنا تقدير الخطأ المعياري، أي S_e والذي يمكن استعماله لإيجاد فترات ثقة واختبار فرضيات حول B, A.

ولكي نتمكن من ذلك نحتاج إلى الفرض السابق وهو أن الأخطاء مستقلة عن بعضها البعض وتخضع لتوزيع طبيعي ، أي أن e_i مستقلة عن e_j لجميع $i \neq j$ وأن e_i تخضع للتوزيع $N(0,\sigma^2)$ جميع i = 1, 2, ... , n.

نظرية (4) :

فترة ثقة $100(1-\alpha)\%$ للمعلمة A هي :

$$\left(a - t\left[1-\frac{\alpha}{2};n-2\right]S_a \;\;,\;\; a + t\left[1-\frac{\alpha}{2};n-2\right]S_a\right)$$

حيث a هي تقدير A في معادلة خط الانحدار، S_a هو تقدير الانحراف المعياري للإحصاء a ويعطى بالمعادلة :

$$S_a^2 = S_e^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)$$

نظرية (5) :

لاختبار الفرضية $H_0 : A = A_0$ مقابل الفرضية البديلة :

(i) $H_1 : A < A_0$

أو (ii) $H_1 : A > A_0$

أو (iii) $H_1 : A \neq A_0$

نستعمل إحصاء الاختبار $t = \frac{a - A_0}{S_a}$ الذي يخضع لتوزيع t ذي (n-2) درجات الحرية.

نظرية (6) :

فترة ثقة $100(1-\alpha)\%$ للمعلمة B هي :

$$\left(b - t\left[1 - \frac{\alpha}{2}; n-2\right] S_b \ , \ b + t\left[1 - \frac{\alpha}{2}; n-2\right] S_b \right)$$

حيث S_b هو تقدير الانحراف المعياري للإحصاء b ويعطى بالمعادلة :

$$S_b^2 = S_e^2 / S_{xx}$$

نظرية (7) :

لاختبار الفرضية $H_0 : B = B_0$ مقابل الفرضية البديلة :

(i) $H_1 : B < B_0$

أو (ii) $H_1 : B > B_0$

أو (iii) $H_1 : B \neq B_0$

نستعمل إحصاء الاختبار $t = \frac{b - B_0}{S_b}$ الذي يخضع لتوزيع t ذي (n-2) درجات الحرية.

مثال (19) : في مثال (17) :

(أ) أوجد فترة ثقة 95% لكل من المعلمين A, B في خط الانحدار .

(ب) اختبر $H_0 : A = 0.8$

مقابل $H_1 : A \neq 0.8$

على مستوى $\alpha = 0.05$.

(جـ) اختبر $H_0 : B = 1.1$

مقابل $H_1 : B \neq 1.1$

على مستوى $\alpha = 0.05$.

الحل : من حل المثال (17) وجدنا b = 1.25 , a = 1 , n = 5

$$S_e = 0.96 \ , \ S_{xx}^2 = 20 \ , \ \bar{x} = 4$$

(أ) $$S_a^2 = S_e^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)$$

$$= 0.9216\left(\frac{1}{5} + \frac{4^2}{20}\right) = 0.9216$$

إذاً $$S_a = 0.96$$

$$t = [0.975;3] = 3.182$$

إذاً فترة ثقة %95 للمعلمة A هي :

$$(1 - 3.182 \times 0.96 \ , \ 1 + 3.182 \times 0.96)$$

أي هي : $(-2.05 , 4.05)$

$$S_b^2 = \frac{S_e^2}{S_{xx}} = (0.96)^2 / 20 = 0.0461$$

إذاً $S_b = \sqrt{0.0461} = 0.215$

فترة ثقة %95 للمعلمة B هي :

$$(1.25 - 3.182 \times 0.215 , \ 1.25 + 3.182 \times 0.215)$$

أي هي : $(0.56 , \ 1.93)$

(ب) (1) اختبر $H_0 : A = 0.8$

(2) مقابل $H_1 : A \neq 0.8$

(3) على مستوى $\alpha = 0.05$.

(4) إحصاء الاختبار :

$t = \frac{a - A_0}{S_\alpha}$ الذي يخضع لتوزيع t ذي (n-2) درجات الحرية.

(5) منطقة الرفض :

ارفض H_0 إذا كان :

$$|t| > t[0.975;3] = 3.182$$

(6) الحسابات :

$$t = \frac{1 - 0.8}{0.96} = 0.208$$

(7) المقارنة : بما أن $0.208 \not> 3.182$ فلا نرفض H_0.

لاحظ أيضا أنه بمقارنة (ب) مع (أ) نجد أن A = 0.8 واقعة ضمن فترة 95% الثقة للمعلمة A ، أي أننا لا نرفض H_0.

(حـ) نتبع الخطوات السابقة ولكن الآن إحصاء الاختبار هو $t = \frac{b - B_0}{S_b}$ يخضع لتوزيع t على درجات حرية (n-2) نحسب قيمة إحصاء الاختبار فنجد :

$$t = \frac{1.25 - 1.1}{0.215} = 0.7$$

وعند المقارنة : $0.7 \not> 3.182$ إذاً لا نرفض H_0.

مثال (20) :

في المثال السابق اختبر $H_0 : A = 4.1$ مقابل $H_1 : A \neq 4.1$ على مستوى دلالة $\alpha = 0.05$.

الحل : نتبع الخطوات السبعة في المثال السابق، والحسابات الآن هي :

$$t = \frac{1 - 4.1}{0.96} = -3.22$$

المقارنة : $3.182 < \left| -3.22 \right|$ إذاً نرفض H_0.

10 : 11 قوة العلاقة الخطية : The Strength of a linear relation

للحصول على مقياس لحسن مطابقة نموذج الخط المستقيم نتفحص نسبة التغير في المتغير التابع التي يمكن عزوها لخط الانحدار. من أجل ذلك ننظر إلى القيمة المشاهدة y_i على أنها مؤلفة من مركبتين، أي :

$$y_i = (a + bx_i) + (y_i - a - bx_i)$$

القيمة المشاهدة | يمكن تفسيرها من خط الانحدار (العلاقة الخطية) | الباقي أو الانحراف عن العلاقة الخطية

إن التغير الكلي في قيم y يعبر عنه بمجموع المربعات S_{yy} الذي تكون SSE جزءا منه، وبالنظر إلى مجموع مربعات البواقي SSE على أنه الابتعاد عن النموذج الخطي ، يتضح أن الفرق بينهما $S_{yy}^2 - SSE$ يمثل الجزء الآخر من التغير.

ولكن ما قيمة $S_{yy}^2 - SSE$ ؟

نحن نعلم أن :

$$SSE = \sum (y_i - a - bx_i)^2 = S_{yy} - \frac{S_{xy}^2}{S_{xx}}$$

إذاً :

$$S_{yy} - SSE = \frac{S_{xy}^2}{S_{xx}}$$

وهذا هو جزء التغير الذي يفسر بالعلاقة الخطية وهو ما نسميه مجموع المربعات للانحدار SSR وإذا عبّرنا عن S_{yy} بالرمز SST أي مجموع المربعات الكلي نحصل على SST = SSR + SSE.

أي أن :

$$S_{yy} = \frac{S_{xy}^2}{S_{xx}} + SSE$$

التغير الكلي في y — **التغير المفسر بالعلاقة الخطية** — **الباقي أو التغير غير المفسر**

وبالرجوع إلى تعريف معامل الارتباط r نجد أن نسبة التغير المفسر بالعلاقة الخطية غلى التغير الكلي

، أي $\frac{S_{xy}^2 / S_{xx}}{S_{yy}}$ هي r^2 وبالتالي نقول :

تعريف (2): تقاس قوة العلاقة الخطية بالمقياس $r^2 = \frac{S_{xy}^2}{S_{xx} S_{yy}}$ التي هي مربع معامل الارتباط ويسمى r^2 معامل التحديد Coefficient of Determination.

مثال (21) : في مثال (17) ما حكمك على صلاحية النموذج الخطي.

الحل : قوة العلاقة الخطية هي :

$$r^2 = \frac{S_{xy}^2}{S_{xx} S_{yy}}$$

ومن جدول حل المثال (17) نجد :

S_{xy} = 145 - 5 × 4 × 6 = 25

S_{xx} = 100 - 5 × 4^2 = 20

S_{yy} = 214 - 5 × 6^2 = 34

إذا $\mathbf{r^2 = \frac{(25)^2}{20 \times 34} = 0.92}$

وتعتبر قوة العلاقة كبيرة، أي أن النموذج صالح.

لاحظ أنه من التعريف (2) $\mathbf{r^2 = \frac{SSR}{SST}}$

أي $\mathbf{r^2 = \frac{SST - SSE}{SST}}$

ولكن $SST = S_{yy} = \sum(y - \bar{y})^2$

أي هي مجموع مربعات الأخطاء لو كنا نتنبأ أو نقدر قيم y بقيمة الوسط الحسابي $\bar{y}$.

ونعلم أن SSE هي $\sum(y - \hat{y})^2$ أي هي مجموع مربعات الأخطاء عندما نتنبأ ونقدر قيم y عن طريق الانحدار وهذا يعني أن r^2 هي نسبة الفرق بين مجموع مربعات الأخطاء في الحالتين إلى المجموع الكلي ولذلك فهي تفسير إلى أي مدى يكون خط الانحدار صالحاً ومناسباً.

تمارين

1-10 : لديك البيانات التالية :

X	1	3	4	2	6	5	8
Y	3	2	5	4	7	6	8

أ- ارسم شكل الانتشار . هل تلاحظ أن هناك علاقة خطية بين Y, X. خمن قيمة لمعامل الارتباط بين Y, X.

ب- احسب r (معامل الارتباط الخطي).

جـ- احسب r_s (معامل ارتباط الرتب).

2-10 : في تمرين (10-1) اختبر الفرضية $\mathbf{H_0 : \rho = 0}$ مقابل $\mathbf{H_1 : \rho \neq 0}$ باستعمال مستوى دلالة $\alpha = 0.05$.

3-10 : يعطي الجدول التالي معدل الطالب في شهادة الدراسة الثانوية X ومعدله التراكمي في الجامعة عند التخرج Y.

X	90.5	87.8	89.2	79.1	81.4
Y	82.0	77.2	76.8	66.2	65.8

أ- ارسم شكل الانتشار.

ب- أوجد معادلة خط انحدار Y على X وارسمه على شكل الانتشار.

جـ- إذا كان معدل طالب في الثانوية العامة 88.5 فكم تتوقع معدله التراكمي عند التخرج.

4-10 : يبين الجدول التالي الطول X (لأقرب سم) والوزن Y (لأقرب كغم) لعشرة طلبة اختيروا عشوائيا من احدى الكليات.

X	175	167	180	178	165	166	182	164	183	170
Y	73	68	75	69	66	65	75	63	74	72

أ- أوجد معادلة انحدار Y على X.

ب- احسب تقديرا للانحراف المعياري للأخطاء أيS_e.

جـ- أوجد فترة ثقة 95% للمعلمة A.

د- أوجد فترة ثقة 95% للمعلمة B.

هـ- علق على مدى حسن مطابقة النموذج الخطي Y = A + Bx + e [أي أوجد r^2].

5-10 : في التمرين (10-4).

أ- اختبر الفرضية $H_0 : A = -100$ مقابل $H_1 : A \neq -100$ على مستوى دلالة 0.05.

ب- اختبر الفرضية $H_0 : B = 0.9$ مقابل $H_1 : B \neq 0.9$ على مستوى دلالة 0.05..

6-10 : في التمرين (10-4) :

أ- احسب معامل الارتباط r.

ب- أوجد فترة ثقة 95% لمعامل الارتباط ρ.

جـ- اختبر الفرضية $H_0 : \rho = 0.8$ مقابل $H_1 : \rho \neq 0.8$ على مستوى دلالة $\alpha =$ 0.05.

7-10 : في التمرين (10-4) احسب معامل الارتباط للرتب r_s.

11-10 : يعطي الجدول التالي عدد الساعات (X) التي يقضيها طالب في الصف العاشر الأساسي في مشاهدة التلفاز في عطلة نهاية الأسبوع وبين علامة الطالب في الرياضيات في بداية الأسبوع التالي (Y).

X	8	9	5	4	11	3	2	6
Y	10	13	15	14	10	16	15	11

أ- احسب معامل الارتباط r.

ب- اختبر الفرضية $H_0 : \rho = 0.8$ مقابل $H_1 : \rho \neq 0.8$ على مستوى دلالة $\alpha =$ 0.05.

جـ- أوجد معادلة خط الانحدار (Y على X).

هل العلاقة طردية أم عكسية ؟

د- علق على مدى صلاحية النموذج الخطي بين Y, X.

12-10 : في تمرين (10-11) :

أ- إذا شاهد الطالب التلفاز لمدة 10 ساعات فكم تتوقع أن يحصل في امتحان الرياضيات.

ب- ما الخطأ في تقديرك لعلامة طالب إذا شاهد التلفاز لمدة 11 ساعة ؟

جـ- اختبر الفرضية $H_0 : A = 15$ مقابل $H_1 : A \neq 15$.

د- اختبر الفرضية $H_0 : B = -0.7$ مقابل $H_1 : B \neq -0.7$ استعمل $\alpha = 0.05$.

الفصل الحادي عشر
تحليل التباين
Analysis of Variance (ANOVA)

1-11 : مقدمة

لقد درسنا في الفصل التاسع طرقا للمقارنة بين معدلي مجتمعين ولكن في كثير من الأحيان نحتاج إلى المقارنة بين عدة مجتمعات. فبينما استعملنا اختبار Z أو اختبار t لاختبار تساوي معدلي مجتمعين معينين فإننا نحتاج إلى طرق عامة بشكل أكبر لنتمكن من اختبار تساوي معدلات ثلاثة مجتمعات أو أكثر. هناك طريقة قوية تسمى تحليل التباين تمكننا من تحليل وتفسير المشاهدات من عدة مجتمعات.

إن من أهم خواص هذه الطريقة الإحصائية أنها تقسم التغير الكلي (Total Variation) في مجموعة من البيانات حسب مصادر التغير الموجودة. ففي حالة المقارنة بين معدلات k من المجتمعات يوجد مصدران للتغير هما :

(1) الفروق بين المعدلات.

(2) التغير ضمن المجتمع الواحد (الخطأ).

2-11 :التصنيف الأحادي (المقارنة بين عدة معاملات - النموذج كامل العشوائية):

One-way Classification (Comparison of Several Treatments -

The Completely Randomized Design).

إن مقارنـة عدة مجتمعات في الوقت ذاتـه يـوفر فـي الوقت والجهـد والتكـاليف التـي نحتاجهـا لـو أجرينـا عدة مقارنـات كل منهـا بـين مجتمعين اثنـين. إن الاصـطلاح "النمـوذج كامـل العشـوائية" مـرادف لأخـذ عينـات عشوائية مستقلة من عدة مجتمعات عندما يعرف المجتمع على أنـه مجتمـع المشاهدات تحت معاملة محددة.

إن أسهل أنواع تحليل التباين هو التصنيف الأحادي وهو عبارة عن تصنيف المشاهدات إلى عدد من المجموعات على أساس خاصية واحدة ويكون ذلك باخضاع عدة تجارب لعدد من المعاملات التي تمثل مستويات الخاصية الواحدة واختبار تأثير هذه المعاملات (Treatments) على المشاهدات التي نحصل عليها.

نعتبر k من المجتمعات ونفترض أنها مستقلة عن بعضها البعض وتخضع للتوزيعات الطبيعية ذات المعدلات $\mu_1, \mu_2, ..., \mu_k$ ونفس التباين σ^2.

نأخذ من المجتمعات المذكورة عينات عشوائية حجم العينة i هو n_i ونرغب في إيجاد طرق مناسبة لاختبار الفرضية :

$$H_0 : \mu_1 = \mu_2 = ... = \mu_k$$

مقابل الفرضية البديلة : يوجد معدلان على الأقل غير متساويين : H_1.

نعبر بالرمز Y_{ij} للمشاهدة ذات الرقم j المأخوذة من المجتمع ذي الرقم i فتظهر المشاهدات كما في الجدول (1).

المجتمع أو المعاملة	المشاهدات				
	1	2	3	...	n
1	Y_{11}	Y_{12}	Y_{13}	...	Y_{1n_1}
2	Y_{21}	Y_{22}	Y_{23}	...	Y_{2n_2}
3	Y_{31}	Y_{33}	Y_{33}	...	Y_{3n_3}

.	.	.	.	...	.
.	.	.	.	...	.
.	.	.	.	...	.
k	Y_{k1}	Y_{k2}	Y_{k3}	...	Y_{kn_k}

الجدول (1)

يمكن كتابة كل مشاهدة على الشكل :

$$Y_{ij} = \mu_i + \varepsilon_{ij}$$

حيث ε_{ij} تقيس انحراف المشاهدة j في العينة i عن معدل المجتمع μ_i .

وباستعمال $\mu_i = \mu + \alpha_i$

حيث

$$\mu = \frac{\sum_{i=1}^{k} \mu_i}{k}$$

يمكن كتابة النموذج أعلاه على الشكل :

النموذج (1). $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$

تحت الشرط :

$$\sum_{i=1}^{k} \alpha_i = 0$$

α_i تعبر عن تأثير المجتمع i (المعاملة i).

أما الافتراضات على النموذج (1) فهي :

ε_{ij} لجميع **i** و **j** تخضع للتوزيع الطبيعي ، مستقلة عن بعضها البعض، ومعدلها **0** ولها نفس التباين σ^2 . وباستعمال النموذج الأخير تصبح الفرضية المبدئية القائلة بتساوي جميع المعدلات مكافئة للفرضية :

$$H_0 : \alpha_1 = \alpha_2 = ... = \alpha_k = 0$$

مقابل الفرضية [واحدة من ألـ α_i على الأقل لا تساوي 0] : H_1.

أما اختبار هذه الفرضية فيبن على المقارنة بين تقديرين مستقلين للتباين σ^2 ونحصل على هذين التقديرين بتقسيم التغير الكلي للبيانات إلى مركبتين، حسب النظرية:

النظرية (1):

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y}_{..}\right)^2 = \sum_{i=1}^{k} n_i\left(Y_{i.} - \overline{Y}_{..}\right)^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y}_{i.}\right)^2$$

حيث $\overline{Y}_{i.} = \dfrac{\sum_{j=1}^{n_i} Y_{ij}}{n_i}$

$$\overline{Y}_{..} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}}{N}$$

ويمكن التعبير عن مجاميع المربعات في النظرية (1) باستعمال الرموز الآتية :

$$1.\ SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y}_{i.}\right)^2$$

وهو مجموع المربعات الكلي .

$$2.\ SSR \sum_{i=1}^{k} n_i\left(Y_{i.} - \overline{Y}_{..}\right)^2$$

وهو مجموع المربعات لمعدلات الصفوف أي مجموع المربعات للمعاملات.

$$3.\ SSE=\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij}-\overline{Y}_{i.}\right)^2$$

وهو مجموع المربعات للخطأ.

وبذلك تصبح النظرية (1).

SST = SSR + SSE

تعتبر SSR مجموع المربعات للمعاملات حيث أن المجتمعات **k** التي أخذت منها العينات يعتبر كل منها مرتبطا بأحدى المعاملات وبذلك تكون المشاهدات Y_{ij} $(j = 1, 2, \ldots, n_i)$ ممثلة n_i من القياسات المرتبطة بالمعاملة**i**.

إن أحد التقديرات المبني على (k - 1) من درجات الحرية، للتباين σ^2 هو

$$S_1^2=\frac{SSR}{k-1}$$

وإذا كانت H_0 صحيحة فإن S_1^2 يكون تقديرا غير منحاز لـ σ^2. أما إذا كانت H_1 صحيحة فإن SSR يميل للكبر وتقدر S_1^2 التباين σ^2 بقيمة أعلى مما يجب.

وهناك تقدير آخر للتباين σ^2 وهو:

$$S_e^2=\frac{SSE}{(N-k)}$$

حيث $N=\sum_{i=1}^{k}n_i$

وهو مستقل عن التقدير الأول، مبني على N-k من درجات الحرية، وغير منحاز في حالة H_0 صحيحة أو غير صحيحة.

أما تباين البيانات المجمعة، أي:

$$S^2=\frac{SSE}{N-k}$$

فهو تقدير لـ σ^2، مبني على (N - 1) من درجات الحرية وغير منحاز في حالة H_0 صحيحة. نلاحظ أن النظرية (1) تقسم درجات الحرية إلى قسمين مرتبطين بمركبتي مجاميع المربعات، أي أن :

$$N - 1 = k - 1 + N - k$$

عندما تكون H_0 صحيحة تكون النسبة :

$$\mathbf{f} = \frac{\mathbf{S_1^2}}{\mathbf{S_e^2}}$$

قيمة من قيم المتغير العشوائي F الذي يخضع لتوزيع F على درجات الحرية (k -1) و (N - k). وبما أن S_1^2 تميل لتقدير σ^2 بقيمة أعلى من الواقع عندما تكون H_0 غير صحيحة، فإننا نستعمل الطرف الأيمن لتوزيع F كمنطقة حرجة لاختبار H_0. أي أن الفرضية H_0 ترفض على مستوى دلالة α إذا كان :

$$\mathbf{f > F[1 - \alpha; k - 1, (N - k)]}$$

ونلخص النتائج السابقة في الجدول (2).

الجدول (2)

مصدر التغير Source of variation	مجموع المربعات Sum of squares	درجات الحرية Degrees of freedom	معدل المربعات Mean square	قيمة f المحسوبة Computed
Due Treatments بسبب المعاملات	SSR	(k - 1)	$S_1^2 = \frac{SSR}{k-1}$	$\frac{S_1^2}{S_e^2}$
Due Error الخطأ	SSE	(n - k)	$S_e^2 = \frac{SSE}{N-k}$	
Total التغير الكلي	SST	(N - 1)		

جدول تحليل التباين الأحادي

One-way ANOVA Table

أما حساب مجموع المربعات SSE , SSR , SST فيكون باستعمال التعابير الصالحة للاستعمال بالآلة الحاسبة والمكافئة للتعاريف الأصلية، فنحسب :

$$T_i = \sum_{j=1}^{n_i} Y_{ij}$$

$$T = \sum_{i=1}^{k} T_i$$

حيث Y_i هو مجموع المشاهدات للمعاملة i ، T مجموع جميع المشاهدات.

$$SST = \sum_{i=1}^{k} \sum_{j=1}^{k} Y_{ij}^2 - \frac{T^2}{N}$$

$$SSR = \sum_{i=1}^{k} \frac{T_i^2}{n_i} - \frac{T^2}{N}$$

SSE = SST - SSR

مثال (1) :

استعملت أربع طرق في تنبيه الطلبة لإرجاع الكتب المستعارة على عينات عشوائية من الطلبة فكانت النتائج بعدد أيام التأخير كما يلي :

طريقة التنبيه	المشاهدات					
I	8	5	9	2		
II	4	3	6	5	2	
III	3	7	4	6	8	8
IV	4	2	3			

هل تدل هذه النتائج أن هناك فروقا في معدلات تأخير إعادة الكتب تعزى لطرق التنبيه ؟

الحل:افرض μ_i = معدل عدد أيام التأخير في الطريقة **i حيث i = 1, 2, 3, 4.**

المطلوب : اختبر $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$.

مقابل الفرضية : يوجد معدّلان على الأقل غير متساويين : H_1

الحسابات : $n_1 = 4, n_2 = 5, n_3 = 6, n_4 = 3, k = 4$

$$T_1 = \sum_{j=1}^{4} T_{1j} = 8+5+9+2 = 24, \overline{Y}_{1.} = \frac{24}{4} = 6$$

$$T_2 = \sum_{j=1}^{5} T_{2j} = 4+3+6+5+2 = 20, \overline{Y}_{2.} = \frac{20}{5} = 4$$

$$T_3 = \sum_{j=1}^{6} T_{3j} = 3+7+4+6+8+8 = 36, \overline{Y}_{3.} = \frac{36}{6} = 6$$

$$T_4 = \sum_{j=1}^{3} T_{4j} = 4+2+3 = 9, \overline{Y}_{4.} = \frac{9}{3} = 3$$

$$T = \sum_{j=1}^{k} T_i = 24+20+36+9 = 89$$

$$N = n_1 + n_2 + n_3 + n_4 = 4+5+6+3 = 18$$

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{ni} Y_{ij}^2 - \frac{T^2}{N}$$

$$= 8^2 + 5^2 + 9^2 + ... + 2^2 + 3^2 - \frac{(89)^2}{18}$$

$$= 531 - 440.06 = 90.94$$

$$SSR = \sum_{i=1}^{k} \frac{T_i^2}{n_i} - \frac{T^2}{N}$$

$$= \frac{(24)^2}{4} + \frac{(20)^2}{5} + \frac{(36)^2}{6} + \frac{(9)^2}{3} - \frac{(89)^2}{18}$$

$$= 467 - 440.06 = 26.94$$

$$SSE = SST - SSR$$

$$= 90.94 - 26.94 = 64$$

درجات الحرية لمجموع المربعات SST هو N - 1 = 18 - 1 = 17

درجات الحرية لمجموع المربعات SSE هو N - k = 18 - 4 = 14

درجات الحرية لمجموع المربعات SSR هو k - 1 = 3

نحسب معدل المربعات لكل مجموع فنجد :

$$\mathbf{S_1^2 = \frac{SSR}{k-1} = \frac{26.94}{3} = 8.98}$$

$$\mathbf{S_e^2 = \frac{SSE}{N-k} = \frac{64}{14} = 4.57}$$

ونحسب f وهي :

$$\mathbf{\frac{S_1^2}{S_e^2} = \frac{8.98}{4.57} = 1.96}$$

نكتب الحسابات في جدول تحليل التباين الأحادي : جدول (4) :

الجدول (4)

مصدر التغير	مجموع المربعات	درجات الحرية	عدل المربعات	f المحسوبة
بسبب المعاملات Due treatments	26.94	3	8.98	1.96
بسبب الخطأ Due error	64	14	4.57	
Total المجموع	90.94	17		

F [0.95 ; 3, 14] = 3.34

وبما أن قيمة f المحسوبة 1.96 أصغر من قيمة F من الجدول 3.34 فلا نرفض H_0 ونعتبر أن ليس هناك فروق ذات دلالة بين معدلات طرق التنبيه المختلفة.

مثال (2) : ابن جدول تحليل التباين الأحادي واختبر الفرضية :

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

للمشاهدات في الجدول التالي :

المعاملة	المشاهدات					
A	9	10	9	8	12	12
B	6	6	9	8	7	6
C	8	7	6	6	3	6
D	7	6	4	5	7	7

الحل : نحسب المقادير المطلوبة في جدول تحليل التباين الأحادي :

$$k = 4, n_1 = n_2 = n_3 = n_4 = 6, N = 24$$
$$T_1 = 9+10+9+8+12+12 = 60 \ ; \ \overline{Y}_{1.} = 10$$
$$T_2 = 6+6+9+8+7+6 = 42 \ ; \ \overline{Y}_{2.} = 7$$
$$T_3 = 8+7+6+6+3+6 = 36 \ ; \overline{Y}_{3.} = 6$$
$$T_4 = 7+6+4+5+7+7 = 36 \ ; \overline{Y}_{4.} = 6$$
$$T = \sum_{1}^{4} T_i = 60+42+36+36 = 174$$
$$N = n_1 + n_2 + n_3 + n_4 = 4 \times 6 = 24$$

$$SST=\sum_{i=1}^{k}\sum_{j=1}^{ni}Y_{ij}^2-\frac{T^2}{N}$$

$$=9^2+10^2+......+7^2+7^2-\frac{174^2}{24}$$

$$=1387-1316.75$$

$$=70.25$$

$$SSR=\sum_{i=1}^{k}\frac{T_i^2}{n_i}-\frac{T^2}{N}$$

$$=\frac{(60)^2}{6}+\frac{(42)^2}{6}+\frac{(36)^2}{6}+\frac{(36)^2}{6}-\frac{(174)^2}{24}$$

$$=1326-1316.75=9.25$$

$$SSE=SST-SSR$$

$$=70.25-9.25=61$$

الجدول (5)

مصدر التغير	مجموع المربعات	درجات الحرية	معدل المربعات	f المحسوبة
بسبب المعاملات SSR	9.25	3	3.08	1.01
بسبب الخطأ SSE	61	20	3.05	
الكلي SST	70.25	23		

جدول التباين الأحادي

اختبار تساوي المعدلات :

1- $H_0:\mu_1=\mu_2=\mu_3=\mu_4$

2- على الأقل إثنان من المعدلات غير متساوية : H_1.

3- $\alpha = 0.05$.

4- إحصاء الاختبار : $f = \frac{S_1^2}{S_2^2}$ يخضع لتوزيع F بدرجات حرية (k-1) , (N-(k).

5- المنطقة الحرجة : نرفض H_0 إذا كان $f > F[0.95 ; 3, 20] = 3.20$.

6- الحسابات : كما ظهرت في الجدول حيث وجدنا $f = 1.01$.

7- الاستنتاج : بما أن $f = 1.01$ أصغر من القيمة الحرجة 3.20 فلا نرفض H_0.

مثال (3) : اختبر الفرضية $H_0 : \mu_1 = \mu_2 = \mu_3$

للمشاهدات في الجدول التالي :

المعاملة	المشاهدات					
A	10	7	6	9		
B	7	8	6	6	5	4
C	5	3	4			

الحل : نحسب المقادير المطلوبة في جدول تحليل التباين الأحادي،

$k = 3$, $n_1 = 4$; $n_2 = 6$; $n_3 = 3$, $N = 4 + 6 + 3 = 13$

$$T_1 = 10 + 7 + 6 + 9 = 32 , T_2 = 7 + 8 + ... + 4 = 36 , T_3 = 5 + 3 + 4 = 12$$

$$\bar{Y}_{1.} = \frac{32}{4} = 8 \ , \ \bar{Y}_{2.} = \frac{36}{6} = 6 \ , \ \bar{Y}_{3.} = \frac{12}{3} = 4$$

$$T = \sum_{1}^{3} T_i = 32 + 36 + 12 = 80$$

$$SST=\sum_{i=1}^{3}\sum_{j=1}^{n_i}Y_{ij}^2-\frac{T^2}{N}$$

$$=10^2+7^2+\ldots\ldots+4^2-\frac{(80)^2}{13}$$

$$=542-492.31$$

$$=49.69$$

$$SSR=\sum_{i=1}^{k}\frac{T_i^2}{n_i}-\frac{T^2}{N}$$

$$=\frac{(32)^2}{4}+\frac{(36)^2}{6}+\frac{(12)^2}{3}-\frac{(80)^2}{13}$$

$$=256+216+48-492.31$$

$$=520-492.31=27.69$$

$$SSE=SST-SSR$$

$$=49.69-27.69=22$$

$$S_1^2=\frac{SSR}{2}=\frac{27.69}{2}=13.84$$

$$S_e^2=\frac{SSE}{N-k}=\frac{22}{13-3}=2.2$$

$$f=\frac{S_1^2}{S_e^2}=\frac{13.84}{22}=6.29$$

نضع الحسابات في جدول تحليل التباين (6) :

الجدول (6)

مصدر التغير	مجموع المربعات	درجات الحرية	معدل المربعات	قيمة f
بسبب المعاملات	27.69	2	13.84	6.29

SSR بسبب الخطأ SSE	22	10	2.2	
الكلي SST	49.69	12		

إجراء الاختبار :

(1) $\mathbf{H_0 : \mu_1 = \mu_2 = \mu_3}$

(2) اثنان من المعدلات على الأقل غير متساوية : H_1.

(3) $\alpha = 0.05$.

(4) إحصاء الاختبار $\mathbf{f = \frac{S_1^2}{S_e^2}}$

يخضع لتوزيع $F(k-1 , N - k)$ أي $F (2, 10)$

(5) منطقة الرفض : ارفض H_0 إذا كان $f > F[0.95 ; 2,10] = 4.10$.

(6) الحسابات : كما موضح في الجدول.

(7) المقارنة : بما أن قيمة $f = 6.29$ أكبر من القيمة الحرجة 4.10 فإننا نرفض H_0 ونستنتج أن اثنين على الأقل من المعدلات $\mathbf{\mu_1, \mu_2, \mu_3}$ غير متساوية.

لاحظ أنه بعد ذكر H_1 , H_0 فإننا لا نحتاج إلى كتابة الخطوات السبعة السابقة ونكتفي بإيجاد النقطة الحرجة ومقارنة قيمة f المحسوبة بها ويكون حلنا كاملا بإعطاء جدول تحليل التباين وهو ما يعطيه الحاسوب في حل مثل هذه المسائل.

مثال (4) :

إنشيء جدول تحليل التباين الأحادي للبيانات التالية : واختبر $\mathbf{H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4}$ على مستوى $\mathbf{\alpha = 0.05}$.

المعاملا	المشاهدات Y_{ij}

ت							
A	1	4	3	2	5		
B	4	8	1	5			
C	7	6	4	5	7	5	8
D	2	5	7	5	6	5	

الحل : باتباع الخطوات كما في الأمثلة السابقة نجد القيم المطلوبة ونضعها في جدول تحليل التباين (7).

الجدول (7)

مصدر التغير	مجموع المربعات	درجات الحرية	معدل المربعات	قيمة f
بسبب المعاملات SSR	26.86	3	8.95	2.64
بسبب الخطأ SSE	61	18	3.39	
الكلي SST	87.86	21		

من جدول توزيع F نجد F [0.95 ; 3,18] = 3.16 إذاً لا نرفض H_0 لأن $\mathbf{3.16 \not< 2.64}$ أي أن f لا تقع في منطقة الرفض.

3-11 : المقارنات المتعددة (Multiple Comparisons)

إن تحليل التباين طريقة قوية لاختبار تساوي عدة معدلات، ولكن إذا رفضنا فرضية تساوي المعدلات وقبلنا الفرضية القائلة بوجود إثنين أو أكثر من المعدلات غير المتساوية فإننا لا نعلم أي من معدلات المجتمعات متساوية وأي منها غير متساوية.

هناك عدة اختبارات تستعمل للإجابة على هذا التساؤل، منها اختبار t المتعدد Multiple test ويسمى اختبار بونفيروني Bonferroni's test وذلك بإجراء اختبار t على الفرق بين كل معدلين على أن يكون مستوى الدلالة الكلي (لجميع الاختبارات معا) هو α.

فإذا ما احتجنا لإجراء m من الاختبارات فإننا نجري هذه الاختبارات على أن يكون مستوى الدلالة لكل منها α/m، وفي هذه الحالة يجب تعيين m مسبقا وقبل إجراء الحسابات. فمثلا إذا كان لدينا 4 مجتمعات (معاملات) وأردنا معرفة أي من المعدلات متساوية وأي منها غير متساوية نأخذ :

$$m = \binom{4}{2} = \frac{4(4-1)}{2} = 6$$

ونجري 6 اختبارات على مستوى الدلالة لكل منها $\alpha/6$.

نجري الاختبار كما يلي :

(1) $H_0 : \mu_i = \mu_j$.

(2) $H_1 : \mu_i \neq \mu_j$.

(3) مستوى الدلالة $\frac{\alpha}{m}$.

(4) إحصاء الاختبار : تحت فرض H_0 صحيحة فإن : $t = \dfrac{\overline{Y}_i - \overline{Y}_j}{S_e\sqrt{\dfrac{1}{n_i} + \dfrac{1}{n_j}}}$

يخضع لتوزيع t ذي درجات حرية (N-k).

(5) المنطقة الحرجة : نرفض H_0 إذا كان $|t| > t\left[1 - \dfrac{\alpha}{2m}; N - k\right]$

(6) الحسابات : احسب t كما في (4).

(7) المقارنة : ارفض H_0 إذا وقعت t في منطقة الرفض (المنطقة الحرجة).

مثال (5) : كانت النتيجة في مثال (3) رفض H_0 أي أن هناك معدلين أو أكثر غير متساوية، أي من هذه المعدلات غير متساوية على مستوى دلالة 0.05 ؟

الحل : نحتاج إلى اختبار الفرضيات الآتية :

$$\mu_1 = \mu_2 \, , \mu_1 = \mu_3 \, , \mu_2 = \mu_3$$

أي أن m = 3 ومستوى الدلالة لكل اختبار يساوي $\frac{0.05}{3} = 0.02$

نجري الاختبار للفرضية الأولى :

(1) $H_0 : \mu_1 = \mu_2$.

(2) $H_1 : \mu_1 \neq \mu_2$.

(3) مستوى الدلالة $\frac{\alpha}{m} = 0.02$.

(4) إحصاء الاختبار : $t = \frac{(\overline{Y}_{1.} - \overline{Y}_{2.}) - (\mu_1 - \mu_2)}{S_e \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

(5) بحساب إحصاء الاختبار تحت H_0 نجد :

$$t = \frac{8 - 6}{\sqrt{2.2}\sqrt{\frac{1}{4} + \frac{1}{6}}} = 2.09$$

حيث استعملنا S_e^2 = 2.2 من جدول تحليل التباين في مثال (3).

(6) المنطقة الحرجة : نرفض H_0 إذا كان $|t| > t[0.99;10] = 2.764$ حيث استعملنا توزيع t على درجات الحرية N - k = 10.

(7) بما أن قيمة t أي 2.09 أصغر من 2.764 لا نرفض H_0 وبالتالي لا يوجد دلالة أن $\mu_1 \neq \mu_2$. بالنظر إلى قيم $\overline{Y}_1, \overline{Y}_2, \overline{Y}_3$ وبحساب إحصاء الاختبار t كما في (4) نجد أن :

t = 3.53 > 2.764 حيث نجد $\mu_1 > \mu_3$ ونستنتج أن $\mu_1 \neq \mu_3$ أما باقي المعدلات فمتساوية أي لا يوجد دلالة أن $\mu_2 \neq \mu_3$.

4-11 : اختبار شفييه Scheffe's Test

إذا كانت نتيجة الاختبار في تحليل التباين عدم رفض H_0 فهذا يعني أن ليس هناك دلالة لرفض أن $\mu_1 = \mu_2 = ... = \mu_k$.

ولكن إذا كانت النتيجة في تحليل التباين رفض الفرضية H_0 فهذا يعني أن هناك بعض المعدلات غير المتساوية وبالتالي نحتاج إلى معرفة أي هذه المعدلات غير متساوية وأيها أكبر من غيرها. لهذا الغرض نجري اختبار شفييه البعدي أي بعد معرفة نتيجة تحليل التباين.

وخلاصة هذا الاختبار أنه تحت شروط وفرضيات التصنيف الأحادي في حالة تساوي حجوم العينات أو عدم تساويها، إذا كان عدد المعاملات يساوي k فإننا نحتاج إلى إجراء المقارنات المتعددة بين كل زوج من المعدلات

$$H_0 : \mu_i - \mu_j = 0$$

مقابل

$$H_1 : \mu_i - \mu_j \neq 0$$

حيث

$$i, j = 1, 2, ..., k, i < j$$

إن التقدير غير المنحاز للفرق $(\mu_i - \mu_j)$ هو $(\overline{Y}_i - \overline{Y}_j)$، وتقدير تباين هذا التقدير هو $S_e^2\left(\frac{1}{n_i} + \frac{1}{n_j}\right)$ حيث S_e^2 هي معدل المربعات للخطأ ، n_j, n_i هما حجما العينتين j , i.

نحسب القيمة f التي تعريفها:

$$f = \frac{\left(\overline{Y}_{i.} - \overline{Y}_{j.}\right)^2}{S_e^2\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

نجد من جدول توزيع F القيمة :

$$F[1-\alpha; k-1, N-k] \quad \text{حيث} \quad N = n_1 + n_2 + ... + n_K$$

ونضرب هذه القيمة بالعدد (k - 1) حيث k عدد المجتمعات (المعاملات).

ضع $S^2 = (k-1)F[1-\alpha; k-1, N-k]$.

نرفض H_0 إذا كان : $f > S^2$

ونعتبر $\mu_1 > \mu_2$ إذا كان :

$$\overline{Y}_i > \overline{Y}_j \quad , \ f > S^2$$

ونعتبر $\mu_1 < \mu_2$ إذا كان :

$$\overline{Y}_i < \overline{Y}_j \quad , \ f > S^2$$

مثال (6) : كانت النتيجة في مثال (3) رفض H_0 أي أن هناك معدلين أو أكثر غير متساوية.
باستعمال اختبار شفييه، أوجد المعدلات غير المتساوية على مستوى دلالة 0.05.

الحل : نختبر أولا :

$$H_0 : \mu_1 - \mu_2 = 0$$

مقابل

$$H_1 : \mu_1 - \mu_2 \neq 0$$

$$\alpha = 0.05$$

$$f = \frac{\left(\overline{Y}_{1.} - \overline{Y}_{2.}\right)^2}{S_e^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \frac{(8-6)^2}{2.2\left(\frac{1}{4} + \frac{1}{6}\right)} = 4.36$$

حيث عوضنا القيم المطلوبة من جدول تحليل التباين.

$$S^2 = (3-1).F[0.95;2;10] = 2 \times 4.10 = 8.20$$

وبما أن $f < S^2$ فلا نرفض H_0.

والآن نجري الاختبار للفرضية :

$$H_0 : \mu_1 - \mu_3 = 0$$

فنحسب

$$f = \frac{\left(\overline{Y}_{i.} - \overline{Y}_{3}\right)^2}{S_e^2\left(\frac{1}{n_1} + \frac{1}{n_3}\right)} = \frac{(8-4)^2}{2.2\left(\frac{1}{4} + \frac{1}{3}\right)} = \frac{16}{1.28} = 12.5$$

وبما أن 8.2 < 12.5 إذن $f > S^2$ ولذلك نرفض H_0. وبما أن $\overline{Y}_{1.} = 8 > \overline{Y}_{3.} = 4$

إذن $\mu_1 > \mu_3$.

وبحساب قيمة f للفرضية الأخيرة :

$$H_0 : \mu_2 - \mu_3 = 0$$

نجد

$$f = \frac{\left(\overline{Y}_{2.} - \overline{Y}_{3.}\right)^2}{S_e^2\left(\frac{1}{n_2} + \frac{1}{n_3}\right)}$$

$$= \frac{(6-4)^2}{2.2\left(\frac{1}{6} + \frac{1}{3}\right)} = \frac{4}{1.1} = 3.64$$

وبما أن 3.64 ليست أكبر من 8.2 أي $f \not> S^2$ فلا نرفض الفرضية H_0 وبالتالي فإن المعدلات غير المتساوية هي μ_3, μ_1.

5-11 : تحليل التباين الثنائي ، نموذج التأثيرات الثابتة :

(Two-way Analysis of Variance, Fixed Effects Model)

يمكن استعمال تحليل التباين لدراسة تأثير عاملين أو أكثر على متغير ما . فمثلا يمكننا دراسة تأثير نوعية القمح ونوعية السماد على إنتاج القمح.

أو دراسة تأثير نوعية المنظفات وطرق استعمالها في نظافة الملابس. أما في الاقتصاد فالأمثلة كثيرة على ضرورة دراسة تأثير عاملين أو أكثر على متغير نود دراسته، فمثلا يمكننا دراسة تأثير مصاريف الدعاية ومناطق البيع على مبيعات بضائع معينة.

ويستعمل تحليل التباين الثنائي كثيرا في التربية لدراسة تأثير طرق التدريس ومستويات الطلبة على التحصيل العلمي، أو دراسة تأثير طرق التدريس وأنواع المدارس على تحصيل الطلبة، وغير ذلك. ويمكن استعمال طرق تحليل التباين في الصناعة، فمثلا يمكن دراسة تأثير نوع العامل المساعد ودرجة حرارة التفاعل على سرعة ذلك التفاعل، هذا بالإضافة إلى الأمثلة الكثيرة في الزراعة وعلم النفس وعلم الاجتماع والمكتبات وغيرها.

في جميع هذه الحالات يمكن استعمال تحليل التباين لدراسة تأثير مستويات كل متغير (كل عامل) على حده بالإضافة إلى دراسة تأثير ائتلاف Combination effect العاملين معا.

سندرس في هذا البند التجارب المؤلفة من عاملين نسمي أحدهما "الصفوف" والآخر " الأعمدة" ونفترض أنهما لا يتفاعلان أي أن تأثير الأعمدة يكون نفسه لكل صف، وأن عدد المشاهدات لكل خلية يساوي 1.

وتسمى مثل هذه التجارب : تجارب الخلايا العشوائية للمقارنة بين r من المعاملات Randomized block experiments for comparing r treatments

حيث في هذه التجارب نضع العناصر المتماثلة في مجموعات متماثلة حجم كل منها r عندما نقارن بين r من المعاملات. بعد ذلك نطبق كل معاملة treatment على عنصر واحد في الخلية وبعدها نجري المقارنات بين استجابات المعاملات من الخلية نفسها. في مثل هذا الحال فإن التغيرات الخارجية يمكن تقليل تأثيرها إلى أكبر حد.

وفي هذه النماذج فإن العشوائية randomization هي حجر الأساس في نماذج الخلايا (block design).

حالما يتم تجميع أفراد التجربة في خلايا، نختار عشوائيا فردا واحدا من الخلية 1 ونطبق عليه المعاملة 1 ثم نختار فردا آخر لنطبق عليه المعاملة الثانية وهكذا حتى يتم تطبيق ألـ r معاملات على ألـ r عناصر في الخلية الأولى، ونعيد هذه العملية على أفراد الخلية الثانية ثم الثالثة وهكذا.

يمكن وضع المشاهدات في جدول (8) كما يأتي :

جدول (8)

الصفوف (Rows)	الأعمدة (Columns)					
	1	2	...	j	...	c
1	Y_{11}	Y_{12}	...	Y_{1j}	...	Y_{1c}
2	Y_{21}	Y_{22}	...	Y_{2j}	...	Y_{2c}
.	.	.		.		.
.	.	.		.		.
.	.	.		.		.
i	Y_{i1}	Y_{i2}		Y_{ij}		Y_{ic}
.	.	.		.		.
.	.	.		.		.
.	.	.		.		.
r	Y_{r1}	Y_{r2}		Y_{rj}		Y_{rc}

وتكتب المشاهدة على الشكل :

$$\mathbf{Y_{ij} = \mu + \alpha_i + \beta_j + \in_{ij}}$$

حيث

$$\mathbf{i = 1,2,...,r}$$
$$\mathbf{j = 1,2,...,c}$$
$$\sum_i \alpha_i = \sum_j \beta_j = 0$$

ε_{ij} لكل i و j مستقلة عن بعضها البعض وتخضع للتوزيع الطبيعي الذي معدله 0 وله نفس التباين σ^2.

r = عدد الصفوف ، c= عدد الأعمدة.

أمـا التغيـر الكلـي فـي تحليـل التبـاين الثنـائي الـذي يفتـرض فيـه عـدم التفاعل بين العاملين فيمكن تقسيمه إلى ثلاث مركبات :

تعزى المركبـة الأولـى للعامـل الأول ، وتعزى الثانيـة للعامـل الثـاني ، وتعزى الثالثة لعامل الصدفة (الخطأ التجريبي) وكل من هذه المركبـات هو عبارة عن مجمـوع مربعات، وهـي فـي حسـابها تشبه تلـك التابعـة للتصنيف الأحادي ونعطي فيما يلي تعريفاتها مع معادلات حسابها بالآلـة الحاسبة :

1- المجموع الكلي للمربعات (Total Sum of Squares)

يقاس تغير المشاهدات بالمجموع الكلي للمربعات SST وهو مجموع مربعات انحرافات المشاهدات عن وسطها الحسابي ويعطى بالمعادلة :

$$SST=\sum_{i=1}^{r}\sum_{j=1}^{c}\left(Y_{ij}-\overline{Y}..\right)^{2}$$

$$=\sum_{i=1}^{r}\sum_{j=1}^{c}Y_{ij}^{\ 2}-rc\overline{Y}^{2}..$$

ودرجات الحرية له (rc-1).

2- مجموع المربعات للصفوف (Sum of Squares for rows)

يقاس التغير بين معدلات الصف بمجموع المربعات للصفوف SSR ويعطى بالمعادلة.

$$SSR=c\sum_{i=1}^{r}\left(Y_{i.}-\overline{Y}..\right)^{2}$$

$$=c\sum_{i=1}^{r}Y_{ij}^{\ 2}-rc\overline{Y}^{2}..$$

ودرجات الحرية له (r - 1).

3- مجموع المربعات للأعمدة (Sum of Squares for Columns)

يقاس التغير بين معدلات العمود بمجموع المربعات للأعمدة SSC بالمعادلة:

$$SSC = r\sum_{j=1}^{c}\left(Y_{.j} - \overline{Y}..\right)^2$$
$$= r\sum_{j=1}^{c} Y_{.j}{}^2 - rc\overline{Y}^2..$$

ودرجات الحرية له (c-1).

4- مجموع المربعات للخطأ (Sum of Squares for Error) :

يقاس مجموع المربعات للخطأ SSE بالتعبير.

$$SSE = \sum_{i=1}^{r}\sum_{j=1}^{c}\left(Y_{ij} - Y_{j.} - Y_{.j} + \overline{Y}..\right)^2$$

ودرجات الحرية له (r-1)(c-1).

نظرية (2) : متطابقة مجموع المربعات

$$\sum_{i=1}^{r}\sum_{j=1}^{c}\left(Y_{ij} - \overline{Y}..\right)^2 = c\sum_{i=1}^{r}\left(Y_{i.} - \overline{Y}..\right)^2 + r\sum_{j=1}^{c}\left(Y_{.j} - \overline{Y}..\right)^2 + \sum_{i=1}^{r}\sum_{j=1}^{c}\left(Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}..\right)^2$$

ويمكن كتابة هذه المتطابقة على الصورة :

SST = SSR + SSC + SSE

ونستعمل هذه المتطابقة لحساب SSE بعد حساب المجموعات الأخرى في 1،2،3.

ويستعمل تحليل التباين الثنائي لاختبار فرضية أن تأثير جميع الصفوف متساو أي أن $\alpha_i = 0$ لكل i = 1, 2, ... , r.

وكذلك فرضية أن تأثير جميع الأعمدة متساو أي أن $\beta_j = 0$، لكل j = 1, 2, ... , c .

ونجري اختبارات هذه الفرضيات كما في تحليل التباين الأحادي ونلخص النتائج في، جدول (9) :

الجدول (9)

مصدر التغير	مجموع المربعات	درجات الحرية	معدل المربعات	قيمة f
بين معدلات الصفوف	SSR	r - 1	$MSR=\frac{SSR}{r-1}$	$\frac{MSR}{MSE}=f_1$
بين معدلات الأعمدة	SSC	c - 1	$MSC=\frac{SSC}{c-1}$	$\frac{MSC}{MSE}=f_2$
الخطأ	SSE	/ \ /	$MSE=\frac{SSE}{(r-1)(c-1)}$	
المجموع الكلي	SST	rc - 1		

تحليل التباين الثنائي - بدون تفاعل - عدد المشاهدات لكل خلية 1.

وتتلخص طريقة اختبار الفرضيات كما يأتي :

لاختبار الفرضية $H_0 : \alpha_i = 0, i = 1,2,...,r$.

أي أن تأثير الصفوف متساو، نجد النقطة الحرجة $F[1-\alpha;(r-1),(r-1)(c-1)]$، ونرفض H_0 على مستوى دلالة α إذا كان : النقطة الحرجة < $\frac{MSR}{MSE}$.

وبنفس الطريقة نختبر الفرضية $H_0 : \beta_i = 0, i = 1,2,...,c$ القائلة بأن تأثير الأعمدة متساو وذلك برفض H_0 على مستوى دلالة α إذا كان :

$$\frac{MSR}{MSE} > F[1-\alpha;(c-1),(r-1)(c-1)]$$

مثال (7) : لتقرير وجود فروق جوهرية تعزى لنوع الحاسوب المستعمل في الطباعة ومهارة السكرتيرة الطابعة أجريت تجربة استعمل فيها ثلاثة حواسيب وأربع سكرتيرات طابعات وسجل عدد الكلمات المطبوعة في الدقيقة فكانت النتائج كما في الجدول التالي :

السكرتيرة الطابعة	الحاسوب		
	I	II	III

A	63	71	61
B	57	65	61
C	56	62	59
D	60	68	64

هل يوجد فروق تعزى لنوع الحاسوب ؟

هل يوجد فروق تعزى للسكرتيرة الطابعة ؟

استعمل مستوى دلالة $\alpha = 0.05$.

الحل : نحسب المعدلات ومجاميع المربعات ونبني جدول تحليل التباين الثنائي.

$$\overline{Y}_{1.} = \frac{63+71+61}{3} = 65 \quad , \quad \overline{Y}_{2.} = \frac{57+65+61}{3} = 61$$

$$\overline{Y}_{3.} = \frac{56+62+59}{3} = 59 \quad , \quad \overline{Y}_{4.} = \frac{60+68+64}{3} = 64$$

$$\overline{Y}_{.1} = \frac{63+57+56+69}{4} = 59$$

$$\overline{Y}_{.2} = \frac{71+65+62+68}{4} = 66.5$$

$$\overline{Y}_{.3} = \frac{61+61+59+64}{4} = 61.25$$

$$\overline{Y}_{..} = \frac{65+61+59+64}{4} = 62.25$$

$$SST=\sum_{i=1}^{4}\sum_{j=1}^{3}Y_{ij}^2-4\times 3\overline{Y}_{..}^2$$
$$=46707-46500.75=206.25$$
$$SSR=3\sum_{i=1}^{4}\overline{Y}_{i.}^2-4\times 3\overline{Y}_{..}^2$$
$$=46569-46500.75=68.25$$
$$SSC=4\sum_{j=1}^{3}\overline{Y}_{.j}^2-4\times 3\overline{Y}_{..}^2$$
$$=46619.25-46500.75=118.5$$
$$SSE=SST-SSR-SSC$$
$$=206.25-68.25-118.5=19.5$$

نعوض المقادير التي حصلنا عليها ونبني جدول تحليل التباين (10).

الجدول (10)

مصدر التغير	مجموع المربعات	**درجات الحرية**	معدل المربعات	قيمة f
بين معدلات الصفوف (بين السكرتيرات)	SSR=68.25	3	22.75	$f_1 = 7$
بين معدلات الأعمدة (بين الحواسيب)	SSC=118.5	2	59.25	$f_2 = 18.23$
الخطأ	SSE = 19.5	6	3.25	
المجموع الكلي	SST=562.25	11		

لاختبار وجود فروق بين السكرتيرات ، نجري اختبار F التالي :

1- نختر الفرضية : $H_0 : \alpha_i = 0, i = 1,2,3,4$ والتي تعني أن معدلات السكرتيرات متساوية.

2- مقابل الفرضية ليس جميع α_i أصفارا : H_1 .

3- $\alpha = 0.05$.

4- إحصاء الاختبر $f_1 = \frac{MSR}{MSE}$ يخضع لتوزيع F بدرجات حرية (r-1),(r-1)(c-1).

5- النقطة الحرجة F[0.95; 3, 6] = 4.76.

6- قيمة f في جدول تحليل التباين هي 7.

7- الاستنتاج : بما أن 4.76 < 7.

نرفض H_0 ونستنتج أن هناك فروقا بين السكرتيرات على مستوى دلالة 0.05.

ولاختبار وجود فروق بين الأنواع المختلفة من الحواسيب، نجري اختبار F التالي :

1- نختبر الفرضية $H_0 : \beta_j = 0, j = 1,2,3$.

2- مقابل الفرضية : ليس جميع β_j أصفارا : H_1 .

3- $\alpha = 0.05$.

4- إحصاء الاختبار $f_2 = \frac{MSC}{MSE}$ يخضع لتوزيع F بدرجات حرية c-1,(r-1)(c-1).

5- النقطة الحرجة F [0.95; 2, 6] = 5.14.

6- قيمة f من جدول تحليل التباين هي 18.23.

7- الاستنتاج : بما أن 5.14 < 18.23.

نرفض H_0 ونستنتج أن هناك فروقا بين الأنواع المختلفة من الحواسيب على مستوى دلالة 0.05.

6-11 تحليل التباين الثنائي مع التفاعل الداخلي ، نموذج التأثيرات الثابتة :

(Two-way Analysis of Variance with Interaction-fixed Effects Model)

ندرس في هذا البند تحليل التباين للتجارب المؤلفة من عاملين نسمي أحدهما "الصفوف" وهو العامل A والآخر "الأعمدة" وهو العامل B ونفترض أنهما يتفاعلان، وأن عدد المشاهدات لكل خلية يساوي n.

يمكن وضع المشاهدات كما يظهر في الجدول (11).

الجدول (11)

الصفوف Rows	الاعمدة Columns				
	1	...	j	...	c
1	Y_{111}		Y_{1j1}		Y_{1c1}
	Y_{112}		Y_{1j2}		Y_{1c2}
	.		.		.
	.		.		.
	.		.		.
	Y_{11n}		Y_{1jn}		Y_{1cn}
.	.		.		.
.	.		.		.
.	.		.		.
i	Y_{i11}		Y_{ij1}		Y_{ic1}
	Y_{i12}		Y_{ij2}		Y_{ic2}
	.		.		.
	.		.		.

	.	.	.
	Y_{i1n}	Y_{ijn}	Y_{icn}
.	.	.	.
.	.	.	.
.	.	.	.
	Y_{r11}	Y_{rj1}	Y_{rc1}
	Y_{r12}	Y_{rj2}	Y_{rc2}
	.	.	.
r	.	.	.
	.	.	.
	Y_{r1n}	Y_{rjn}	Y_{rcn}

(المشاهدات في تحليل التباين الثنائي، عدد المشاهدات في الخلية يساوي n).

وكما في حالة تحليل التباين الأحادي يمكننا كتابة نموذج تحليل التباين الثنائي على الشكل:

$$\mathbf{Y_{ijk}} = \mu_{ij} + \in_{ijk}$$

حيث μ_{ij} هي المعلمات ، $\in_{ijk}$ مستقلة عن بعضها البعض وتخضع للتوزيع $N(0, \sigma^2)$.

$i = 1, 2, ... r ; \quad j = 1, 2, ... c ; \quad k = 1, 2, ... n$

إن المعلمة μ_{ij} هي معدل الاستجابة للمعاملة التي يكون فيها العامل A على المستوى i، والعامل B على المستوى j . وبكتابة $\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$.

ثم تعويض μ_{ij} في نموذج تحليل التباين الثنائي نحصل على النموذج:

$$\mathbf{Y_{ijk}} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \in_{ijk}$$

حيث :

$i = 1, 2, \dots r \quad j = 1, 2, \dots c \quad k = 1, 2, \dots n$

وتحت الشروط :

$$\sum_i \alpha_i = \sum_j \beta_j = 0$$
$$\sum_i (\alpha\beta)_{ij} = 0$$
$$\sum_j (\alpha\beta)_{ij} = 0$$

$\in_{ijk}$ لكل i و j و k مستقلة عن بعضها البعض وتخضع للتوزيع الطبيعي الذي معدله 0 وله نفس التباين σ^2.

r = عدد الصفوف ، c = عدد الأعمدة ، n = عدد المشاهدات لكل خلية.

ولكي نتمكن من إظهار التركيب العاملي الذي يظهر تأثير العامل A والعامل B فإننا نكتب :

$$\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$$

وهذا يعني أن μ_{ij} تقسم إلى أربعة أقسام :

1- ثابت عام هو μ وهو المعدل العام ويساوي $\sum_i \sum_j \mu_{ij} / rc$

2- التأثير الأساسي α_i للعامل A على المستوى i .

3- التأثير الأساسي β_j للعامل B على المستوى j .

4- تأثير التفاعل التداخلي $(\alpha\beta)_{ij}$ عندما يكون العامل A على المستوى i والعامل B على المستوى j.

ويستعمل تحليل التباين الثنائي ذي التفاعل، لاختبار فرضية أن تأثير جميع الصفوف متساو (تأثير جميع مستويات العامل A متساو) أي أن :

$\alpha_i = 0$ لجميع $i = 1, 2, \dots. r$

مقابل الفرضية البديلة : ليس جميع α_i مساوية للصفر.

وكذلك فرضية أن تأثير جميع الأعمدة متساو (تأثير جميع مستويات العامل B متساوية) أي أن $\beta_i =$ 0لجميع $j = 1, 2, \dots c$.

مقابل الفرضية البديلة : ليس جميع β_j مساوية للصفر.

وأيضا يستعمل هذا التحليل لاختبار فرضية عدم وجود تفاعل بين الصفوف والأعمدة (عدم وجود تفاعل بين مستويات العاملين B , A) أي أن : $(\alpha\beta)_{ij} = 0$ لجميع $j = 1, 2, \dots c, i = 1, 2, \dots r$.

مقابل الفرضية البديلة ليس جميع $(\alpha\beta)_{ij}$ مساوية للصفر.

نجري اختبارات هذه الفرضيات كما في تحليل التباين الثنائي (بدون تفاعل) ونلخص النتائج في الجدول (12).

الجدول (12)

مصدر التغير	مجموع المربعات	درجات الحرية	معدل المربعات	قيمة f
بين معدلات الصفوف (العامل A)	SSR	(r - 1)	$MSR = \frac{SSR}{r-1}$	$\frac{MSR}{MSE}$
بين معدلات الأعمدة (العامل B)	SSC	(c - 1)	$MSC = \frac{SSC}{c-1}$	$\frac{MSC}{MSE}$
التفاعل بين الصفوف والأعمدة	SSRC	$(r-1)(c-1)$	$MSE = \frac{SSE}{(r-1)(c-1)}$	$\frac{MSRC}{MSE}$
الخطأ	SSE	$rc(n-1)$	$MSE = \frac{SSE}{rc(n-1)}$	
المجموع الكلي	SST	nrc - 1		

تحليل التباين الثنائي ذي التفاعل.

تقسيم مجموع المربعات الكلي :

يمكن تقسيم التغير الكلي في تحليل التباين الثنائي ذي التفاعل إلى أربع مركبات وذلك بتقسيم انحراف المشاهدة Y_{ijk} عن المعدل العام ... $\overline{Y}$ إلى أربعة أقسام ثم تربيع الطرفين وإيجاد حاصل الجمع لجميع المشاهدات مع ملاحظة أن جمع حواصل الضرب التصالبي يساوي صفرا.

$$\left(Y_{ijk}-\overline{Y}_{...}\right)=\left(\overline{Y}_{i..}-\overline{Y}_{...}\right)+\left(\overline{Y}_{.j.}-\overline{Y}_{...}\right)+\left(\overline{Y}_{ij.}-\overline{Y}_{i..}-\overline{Y}_{.j.}+\overline{Y}_{...}\right)+\left(Y_{ijk}-\overline{Y}_{ij}.\right)$$

الانحراف حول معدل المعاملة + تأثير التفاعل AB + التأثير الأساسي للعامل B + التأثير الأساسي للعامل A = الانحراف الكلي.

أما مجموع المربعات الكلي فهو :

$$SST=\sum_i\sum_j\sum_k\left(Y_{ijk}-\overline{Y}_{...}\right)^2$$

ودرجات الحرية له تساوي (nrc-1).

ومركبات SST هي:

1. مجموع مربعات العامل A (الصفوف): يقاس التغير بين تقديرات معدلات العامل A (الصفوف) بمجموع المربعات للصفوف.

$$SSR=\sum_i\sum_j\sum_k\left(\overline{Y}_{i..}-\overline{Y}_{...}\right)^2$$
$$=nc\sum_i(\overline{Y}_{i..}-\overline{Y}_{...})^2$$

ودرجات الحرية له (r - 1).

2- مجموع مربعات العامل B (الأعمدة) : يقاس التغير بين تقديرات معدلات العامل B (الأعمدة) بمجموع المربعات للأعمدة.

$$SSC=\sum\sum\sum\left(Y_{ijk}-\overline{Y}_{...}\right)^2$$

$$=nrc\sum_j\left(\overline{Y}_{.j.}-\overline{Y}_{...}\right)^2$$

ودرجات الحرية له تساوي (c-1).

3- مجموع مربعات التفاعل AB (التفاعل التداخلي بين الصفوف والأعمدة) : يقاس التغير بين تقديرات التفاعلات للمعاملات التي عددها rc بمجموع المربعات للتفاعل.

$$\mathbf{SSRC} = \sum_i \sum_j \sum_k \left(\overline{\mathbf{Y}}_{ij.} - \overline{\mathbf{Y}}_{i..} - \overline{\mathbf{Y}}_{.j.} + \overline{\mathbf{Y}}_{...}\right)^2$$

$$= \mathbf{n} \sum_i \sum_j \left(\overline{\mathbf{Y}}_{ij.} - \overline{\mathbf{Y}}_{i..} - \overline{\mathbf{Y}}_{.j.} + \overline{\mathbf{Y}}_{...}\right)^2$$

ودرجات الحرية له تساوي (c-1) (r-1).

4- مجموع مربعات الخطأ : وهو الذي يقيس التغير بين المعاملات، وهو :

$$\mathbf{SSE} = \sum_i \sum_j \sum_k \left(\overline{\mathbf{Y}}_{ijk} - \overline{\mathbf{Y}}_{ij.}\right)^2$$

ودرجات الحرية له تساوي rc(n-1).

هذه تعاريف مجاميع المربعات أما حساب كل منها فيعطى بالنظرية التالية :

نظرية (3) : يقسم مجموع المربعات الكلي إلى أربع مركبات، نضعها بالمتطابقة التالية ، حيث الرموز كما ذكرنا آنفا :

$$SST = SSR + SSC + SSRC + SSE$$

أما المعادلات التي تستعمل لحسابات هذه المجاميع فهي :

$$\mathbf{SST} = \sum_i \sum_j \sum_k \overline{\mathbf{Y}}^2_{ijk} - \mathbf{nrc}\overline{\mathbf{Y}}^2_{...}$$

$$\mathbf{SSR} = \mathbf{nc} \sum_i \overline{\mathbf{Y}}^2_{i..} - \mathbf{nrc}\overline{\mathbf{Y}}^2_{...}$$

$$\mathbf{SSC} = \mathbf{nr} \sum_j \overline{\mathbf{Y}}^2_{.j.} - \mathbf{nrc}\overline{\mathbf{Y}}^2_{...}$$

$$\mathbf{SSE} = \sum_i \sum_j \sum_k \overline{\mathbf{Y}}^2_{ijk} - \mathbf{n} \sum_i \sum_j \overline{\mathbf{Y}}^2_{ij.}$$

$$\mathbf{SSRC} = \mathbf{SST} - \mathbf{SSR} - \mathbf{SSC} - \mathbf{SSE}$$

ونجري اختبارات الفرضيات السابقة كما يلي :

(i) لاختبار الفرضية :

$$H_0 : \alpha_i = 0 \quad i = 1, 2, \dots, r$$

مقابل الفرضية : ليس جميع α_i مساوية للصفر : H_1

نجد النقطة الحرجة $C_1 = F[1-\alpha; r-1.rc(n-1)]$ ونرفض H_0 على مستوى دلالة α إذا كان :

$$f_1 = \frac{MSR}{MSE} > C_1$$

(ii) وبنفس الطريقة ، نختبر الفرضية :

$H_0 : \beta_j = 0 \quad j = 1, 2, \dots, c$ على مستوى دلالة α.

مقابل الفرضية : ليس جميع β_j مساوية للصفر : H_1

فنرفض H_0 إذا كان :

$$f_2 = \frac{MSC}{MSE} > C_2 = F[1-\alpha; c-1, rc(n-1)]$$

(iii) وكذلك نختبر الفرضية :

$H_0 : (\alpha\beta)_{ij} = 0,\ i = 1, 2, \dots, r ; j = 1, 2, \dots, c$ على مستوى دلالة α.

مقابل الفرضية : ليس جميع $(\alpha\beta)_{ij}$ مساوية للصفر : H_1.

بأن نرفض H_0 إذا كان :

$$f_3 = \frac{MSRC}{MSE} > C_3 = F[1-\alpha; (r-1)(c-1), rc(n-1)]$$

مثال (8) :

وضع 72 طالبا من الصف السادس الأساسي في تسع شعب عشوائيا

وقام ثلاثة معلمين بتدريس هذه الشعب بثلاث طرق تدريس مختلفة (كل معلم درس ثلاث شعب بطريقة مختلفة لكل شعبة) وفي نهاية التجربة التي دامت شهرا عقد للطلبة امتحان في موضوع الرياضيات الذي درسوه خلال الشهر فكانت النتائج كما يلي (الامتحان من 20) :

المعلم	طريقة التدريس					
	I		II		III	
(1)	12	18	11	16	13	11
	17	17	10	12	9	12
	14	15	8	13	8	14
	16	11	14	12	10	11
(2)	19	18	9	11	13	14
	15	14	12	10	15	10
	17	16	19	9	11	16
	16	13	8	11	12	13
(3)	18	14	15	13	10	14
	16	12	14	16	12	11
	20	18	12	11	15	13
	15	15	17	14	9	12

أ- اكتب جدول تحليل التباين.

ب- اختبر الفرضية : لا يوجد فروق ذات دلالة على مستوى 0.05 في تحصيل الطلبة تعزى للمعلم.

جـ- اختبر الفرضية: لا يوجد فروق ذات دلالة على مستوى 0.05 في تحصيل الطلبة تعزى لطريقة التدريس.

د- اختبر الفرضية : لا يوجد فروق ذات دلالة عل مستوى 0.05 في تحصيل الطلبة تعزى لطريقة التدريس والمعلم.

الحل : نحسب المعدلات المختلفة من المعادلات الآتية :

$$\overline{Y}_{ij.} = \frac{\sum_k Y_{ijk}}{n} \qquad \overline{Y}_{.j.} = \frac{\sum_i \sum_k Y_{ijk}}{rn}$$

$$\overline{Y}_{i..} = \frac{\sum_j \sum_k Y_{ijk}}{cn} \qquad \overline{Y}_{...} = \frac{\sum_i \sum_j \sum_k Y_{ijk}}{rcn}$$

$\overline{Y}_{11.} = 15$ $\qquad \overline{Y}_{21.} = 16$ $\qquad \overline{Y}_{31.} = 16$

$\overline{Y}_{12.} = 12$ $\qquad \overline{Y}_{22.} = 10$ $\qquad \overline{Y}_{32.} = 14$

$\overline{Y}_{13.} = 11$ $\qquad \overline{Y}_{23.} = 13$ $\qquad \overline{Y}_{33.} = 12$

$\overline{Y}_{1..} = 12.67$ $\qquad \overline{Y}_{2..} = 13$ $\qquad \overline{Y}_{3..} = 14$

$\overline{Y}_{.1.} = 15.67$ $\qquad \overline{Y}_{2..} = 12$ $\qquad \overline{Y}_{3..} = 12$

$\overline{Y}_{...} = 13.22$

نحسب مجاميع المربعات كما جاءت في النظرية :

$$SST = \sum_i \sum_j \sum_k Y_{ijk}^2 - rcn\overline{Y}_{...}^2$$

$$= 13172 - 12583.325 = 588.68$$

$$SSE = \sum_i \sum_j \sum_k Y_{ijk}^2 - n\sum_i \sum_j \overline{Y}_{ij.}^2$$

$$= 13172 - 8 \times 1611 = 284$$

$$SSR = cn\sum_i \overline{Y}_{i..}^2 - rcn\overline{Y}_{...}^2$$

$$= 13 \times 8 \times 525.529 - 3 \times 3 \times 8 \times 174.78$$

$$= 112612.69 - 12583.33 = 29.37$$

$$SSC = rn\sum \overline{Y}_{.j.}^{2} - rcn\overline{Y}_{...}^{2}$$

$$= 3 \times 8 \times 533.549 - 12583.33$$

$$= 12805.17 - 12583.33$$

$$= 221.84$$

$$SSRC = SST - SSR - SSC - SSE$$

$$= 588.68 - 29.37 - 221.84 - 284$$

$$= 53.47$$

(أ) بعد إجراء الحسابات كلها نلخص النتائج في جدول تحليل التباين (13).

الجدول (13)

مصدر التغير	مجموع المربعات	درجات الحرية	معدل المربعات	قيمة f
بين معدلات الصفوف (عامل المعلمين)	29.37	2	14.685	$3.26 = f_1$
بين معدلات الأعمدة (عامل طرق التدريس)	221.84	2	110.92	$24.59 = f_2$
التفاعل بين المعلمين وطرق التدريس	53.47	4	13.37	$2.96 = f_3$
الخطأ	284	63	4.51	
المجموع الكلي	588.68	71		

(ب) من جدول F ، النقطة الحرجة هي :

$F\ [0.95\ ;\ 2,\ 63] = 3.15$

وبما أن $f_1 = 3.26$ أكبر من 3.15 إذاً نرفض H_0 ، أي أن هناك فروق في تحصيل الطلبة ذات دلالة 0.05 تعزى للمعلم.

(جـ) كما في ب، $F\ [0.95\ ;\ 2, 63] = 3.15$ وبما أن $f_2 = 24.59$ أكبر من 3.15 إذاً نرفض H_0، أي أن هناك فروق في تحصيل الطلبة على مستوى 0.05 تعزى لطرق التدريس.

(د) $F\ [0.95\ ;\ 4, 63] = 2.53$

بما أن $f_3 = 2.96$ أكبر من 2.53 إذاً نرفض H_0 ونستنتج أن هنا ك فروقا في تحصيل الطلبة على مستوى دلالة 0.05 تعزى للتفاعل بين المعلم وطريقة التدريس. لاحظ أن جداول F غير موجودة لدرجات الحرية في المقام 63 ولذلك تأخذ أقرب قيمة وهي المقابلة لدرجات الحرية 60.

7-11 : اختبارات شفييه وبونفيروني البعدية

عندما يؤدي اختبار F إلى رفض الفرضية H_0، المتعلقة بالعامل A ، فإننا نحتاج إلى معرفة أي المستويات من العامل A (الصفوف) مختلفة عن بعضها البعض وأي منها متساو، وبالمثل، إذا رفضت الفرضية H_0 المتعلقة بالعامل B (الأعمدة) فإننا نحتاج إلى معرفة أي المستويات من العامل B متساوية وأي منها مختلفة عن بعضها البعض.

إن اختبارات المقارنات المتعددة التي بحثناها في التصنيف الأحادي تبقى صالحة لاختبار تساوي المستويات ومعرفة أي منها مختلف عن الآخر شريطة استعمال التقديرات المناسبة ودرجات الحرية المناسبة كما يلي :

(1) اختبار شفييه :

اختبار شفييه المتعدد لاختبار الفروق بين معدلات الصفوف نستعمل :

$$S_r^2 = (r-1)F[1-\alpha;(r-1), rc(n-1)]$$

فعند اختبار $H_0 : \alpha_i - \alpha_{i'} = 0$ لجميع $i < i'$ فهذا يعني اختبار :

$$H_0 : (\mu + \alpha_i) - (\mu + \alpha_{i'}) = 0$$

نرفض H_0 على مستوى دلالة α إذا كان :

$$\frac{\left(\overline{Y}_{i..}-\overline{Y}_{i'..}\right)^2}{S_e^2\frac{2}{cn}}>S_r^2$$

أما الاختبار للفروق بين معدلات الأعمدة فنستعمل :

$$S_c^2=(c-1)F[1-\alpha;(c-1),rc(n-1)]$$

وعند اختبار $H_0:\beta_j-\beta_{j'}=0$ لجميع $j<j'$ فهذا يعني اختبار

$$H_0:(\mu+\beta_j)-(\mu+\beta_{j'})=0$$

نرفض H_0 على مستوى دلالة α إذا كان :

$$\frac{\left(\overline{Y}_{.j.}-\overline{Y}_{..j'}\right)^2}{S_e^2\frac{2}{rn}}>S_c^2$$

(2) اختبار بونفيروني أو t المتعدد :

(i) اختبار t المتعدد لاختبار الفروق بين معدلات الصفوف مأخوذة اثنين اثنين، نعين أولا عدد الاختبارات المطلوبة m حيث :

$$m=\frac{r(r-1)}{2}$$

نستعمل مستوى الدلالة $\frac{\alpha}{m}$ ونجري اختبار الفرضية الأولى : $\alpha_i=\alpha_{i'}$ أي :

(1) $H_0:\mu+\alpha_i=\mu+\alpha_{i'}$

(2) مقابل $H_1:\mu+\alpha_i\neq\mu+\alpha_{i'}$

(3) مستوى الدلالة $\frac{\alpha}{m}$.

(4) إحصاء الاختبار : تحت فرض H_0 صحيحة فإن :

$$t = \frac{\overline{Y}_{i..} - \overline{Y}_{i'..}}{S_e\sqrt{\frac{2}{cn}}}$$

يخضع لتوزيع t ذي درجات الحرية rc (n-1). حيث S^2_e هو معدل مربعات الخطأ.

(5) منطقة الرفض : أرفض : H_0 إذا كان $|t| > t\left[1-\frac{\alpha}{2m}; rc(n-1)\right]$.

(6) الحسابات : احسب t في (4).

(7) المقارنة : قارن t المحسوبة مع t الجدولية أي هل تقع t المحسوبة داخل منطقة الرفض وعندها ارفض H_0.

(ii) اختبار t المتعدد لاختبار الفروق بين معدلات الأعمدة مأخوذة اثنين اثنين.

نجري الخطوات نفسها في (i) إلا أن إحصاء الاختبار يصبح :

$$t = \frac{\overline{Y}_{.j.} - \overline{Y}_{.j'.}}{S_e\sqrt{\frac{2}{rn}}}$$

مثال (9) :

في مثال (8) رفضنا الفرضية في (ب)، أوجد أي الفروق ذات دلالة على مستوى $\alpha = 0.05$ باستعمال اختبار شفييه ثم اختبار بونفيروني (t المتعدد).

الحل :

أ. اختبار شفييه: r = 3 فنحتاج إلى اختبار $\alpha_1 = \alpha_2$ ، $\alpha_1 = \alpha_3$, $\alpha_2 = \alpha_3$

نختبر الفرضية الأولى $H_0 : \alpha_1 = \alpha_2$ أي $H_0 : \mu + \alpha_1 = \mu + \alpha_2$ مقابل $H_1 : \mu + \alpha_1 \neq \mu + \alpha_2$ مستوى الدلالة 0.05.

$$S_r^2 = (3-1)F[0.95;2,63] = 2 \times 3.15 = 6.3$$

$$S^2 = \frac{(\overline{Y}_{1..} - \overline{Y}_{2..})}{S_e^2\frac{2}{cn}} = \frac{(12.67-13)}{4.51 \times \frac{2}{3\times 8}} = 0.29$$

وبما أن $6.3 \ngtr 0.29$ لا نرفض H_0 أي أنه لا يوجد فروق جوهرية بين المعلم (1) والمعلم (2).

الآن ، نكرر حساب S^2 بين المعلم (1) والمعلم (3) :

$$S^2 = \frac{(12.67-14)^2}{4.51 \times \frac{2}{3 \times 8}} = 4.7$$

وبما أن $6.3 \ngtr 4.7$ لا نرفض H_0.

وأخيرا نجري حساب S^2 بين المعلم (2) والمعلم (3) :

$$S^2 = \frac{(13-14)^2}{4.51 \times \frac{2}{3 \times 8}} = 2.66$$

وبما أن $6.3 \ngtr 2.66$ لا نرفض H_0.

ب . اختبار t المتعدد :

بما أن r= 3 فإن $m = \frac{3 \times 2}{2} = 3$ إذاً مستوى الدلالة تقريباً $0.02 = \frac{0.05}{3}$. لاختبار
$H_0 : \alpha_1 = \alpha_2$ أي $H_0 : \mu + \alpha_1 = \mu + \alpha_2$ نحسب إحصاء الاختبار :

$$t = \frac{(\overline{Y}_{1..} - \overline{Y}_{2..})}{S_e \sqrt{\frac{2}{cn}}} = \frac{12.67-13}{\sqrt{4.51}\sqrt{\frac{2}{3 \times 8}}} = -0.47$$
$$t[0.99;63] = 2.358$$

وبما أن $2.358 \ngtr |-0.47|$ فلا نرفض H_0.

الآن نختبر $H_0 : \alpha_1 = \alpha_3$ نحسب :

$$t = \frac{(\overline{Y}_{1..} - \overline{Y}_{2..})}{S_e\sqrt{\frac{2}{cn}}} = \frac{|12.67 - 14|}{\sqrt{4.51}\sqrt{\frac{2}{3\times 8}}} = \frac{-4.61}{2.12} = -2.17$$

لا نرفض H_0.

وأخيرا $H_0 : \alpha_2 = \alpha_3$ لا نرفضها أيضا.

لاحظ أنه بالرغم من رفض فرضية تساوي المعدلات في مثال (8) إلا أن اختبار شفييه واختبار t المتعدد لم تكشف أي من المعدلات غير متساوية. إن السبب في ذلك هو قرب قيمة f (f = 3.26) من القيمة الجدولية 3.15، هذا من جهة ومن جهة أخرى استعملنا $\alpha = 0.05$ ولو استعملنا مستوى دلالة أكبر مثلا $\alpha = 0.10$ لوجدنا في اختبار t المتعدد :

$$t\left[1-\frac{\alpha}{2m};rc(n-1)\right] =$$
$$t\left[1-\frac{0.10}{6};63\right] = t[0.98;63] = 2.05$$

فنجد لاختبار $H_0 : \alpha_1 = \alpha_3$ أن $|t| = |-2.17| > 2.05$.

إذا نرفض H_0 ونستنتج أن معدل التحصيل لتدريس المعلم الأول يختلف عن المعلم الثالث وذلك لصالح المعلم الثالث (معدل التحصيل لتدريس المعلم الثالث 14، وللمعلم الأول 12.67).

وبالنسبة لاختبار شفييه، نجد:

$$S_r^2 = (3-1)F[0.90;2,63] = 2\times 2.39 = 4.78$$

وكما وجدنا $S^2 = 4.7$ بين المعلم (1) والمعلم (3) ولما كان $4.7 \not> 4.78$ (مع أنها قريبة منها كثيرا) فلا نرفض H_0.

من هذا نستنتج أن اختبار t المتعدد يمكن أن يرفض الفرضية بينما لا يرفضها اختبار شفييه.

مثال (10) : كانت النتيجة في مثال (8) رفض تساوي المعدلات (أي وجود فروق) تعزى لطريقة التدريس. أوجد أي من طرق التدريس غير متساوية.

الحل : نعيد الاختبارات كما في مثال (9) ولكن الآن نختبر الفرضيات :
$\beta_2=\beta_3\ ,\beta_1=\beta_3\ ,\beta_1=\beta_2$ باستعمال $\alpha=0.05$.

(1) لاختبار الفرضية $\beta_1=\beta_3$ (اختبار شفييه) نجد :

$$S_c^2=(c-1)F[0.95;c-1,rc(n-1)]$$
$$=2\times 3.15=6.30$$

$$S^2=\frac{\left(\overline{Y}_{.1.}-\overline{Y}_{.2.}\right)^2}{S_e^2\dfrac{2}{rn}}=\frac{(15.67-12)^2}{4.51\times\dfrac{2}{3\times 8}}=35.83$$

وبما أن $S^2>S_c^2$ حيث (35.83 > 6.30) نرفض H_0 أي أن طريقة التدريس (I) مختلفة عن (II) ونستنتج أن طريقة التدريس (I) تعطى معدلا أعلى من طريقة التدريس (II) وذلك لأن $\overline{Y}_{.1.}>\overline{Y}_{.2.}$.

ولاختبار الفرضيات الأخرى ، نلاحظ أن $\overline{Y}_{.1.}-\overline{Y}_{.3.}=15.67-12$ ، إذاً فقيمة S^2 لاختبار $H_0:\beta_1=\beta_3$ تبقى نفسها 35.83 ونستنتج أن طريقة التدريس (I) أفضل من الطريقة (III) أما $H_0:\beta_2=\beta_3$ فلا نرفضها لأن $\overline{Y}_{.2.}=\overline{Y}_{.3.}$.

(2) باستعمال t المتعدد ، لاختبار $H_0:\beta_1=\beta_2$ نحسب

$$t=\frac{\left(\overline{Y}_{.1.}-\overline{Y}_{.2.}\right)}{S_e\sqrt{\dfrac{2}{cn}}}$$
$$=\frac{15.67-12}{\sqrt{4.51}\sqrt{\dfrac{2}{3\times 8}}}=5.99$$

ونجد من الجدول :

$$t\left[1-\frac{\alpha}{2m};rc(n-1)\right]=t[0.99;63]=2.358$$

وبالمقارنة فإن 2.358 < 5.99 ، إذاً نرفض H_0 إذاً طريقة التدريس (I) تعطي معدلاً أعلى من طريقة التدريس II وبما أن $\overline{Y}_{.2.} = \overline{Y}_{.3.}$ نجد t لاختبار $H_0 : \beta_1 = \beta_3$ هي 5.99 فنرفض هذه الفرضية ولكن لا نرفض $H_0 : \beta_2 = \beta_3$.

تمارين

1-11 : درست ثلاث شعب من الصف الرابع الأساسي من قبل ثلاثة معلمين فكانت النتائج في الرياضيات كما يأتي :

المعلم	العلامات									
A	91	82	77	53	76	65	54	86	48	93
B	55	65	80	61	74	68	67	40	68	67
C	55	88	78	90	42	87	73	75	50	86

أ- أجر تحليل التباين. هل هناك تفاوت في معدلات الشعب المختلفة ؟ استعمل مستوى دلالة 0.05.

ب- استعمل اختبار t المتعدد للمقارنة بين معدلات الشعب المختلفة. استعمل مستوى دلالة 0.05.

جـ- استعمل اختبار شفييه للإجابة عن فرع (ب).

2-11 : أربع شعب في مساق إحصاء درّست من قبل أربعة مدرسين فكانت علاماتهم النهائية كما يأتي :

المدرس			
A	B	C	D
95	70	40	85
72	95	80	70
65	40	90	78
50	85	75	95
40	60	80	40
75	90	85	52

69	55	95	67
65	65	50	
70	60	62	
91	70		
55			
90			
70			
75			
68			

أ- هل يوجد اختلاف ذو دلالة بين المعدلات لشعب المدرسين الأربعة. استعمل مستوى دلالة 0.05.

ب- استعمل اختبار t المتعدد للمقارنة بين معدلات الشعب الأربع.

3-11 : في تحليل التباين الأحادي ، اشتق المعادلة الخاصة بحساب كل من SST و SSR.

4-11 : في تحليل التباين الثنائي بمشاهدة واحدة لكل خلية، اشتق المعادلة الخاصة بحساب كل من SST , SSR ، SSC.

5-11 : استعملت أربعة أنواع من الأسمدة على ثلاثة أنواع من القمح فأعطت القراءات التالية :

نوع السماد	نوع القمح		
	A	B	C
I	75	73	79
	68	68	72
	73	70	76

	74 70	73 71	78 67
II	68 69 67 73 68	75 73 70 69 68	74 72 68 69 72
III	60 63 63 68 66	60 58 62 67 63	58 57 60 61 59
IV	70 65 56 44 60	72 63 67 70 63	71 70 72 58 74

أ- إبنِ جدول تحليل التباين.

ب- على افتراض أن قطع الأرض لها نفس الخصوبة، اختبر فرضية أن أنواع القمح لها نفس التأثير على الإنتاج : على مستوى دلالة $\alpha = 0.05$.

جـ- اختبر فرضية أن أنواع الأسمدة لها نفس التأثير على الإنتاج. استعمل مستوى دلالة α = 0.05.

د- اختبر فرضية عدم وجود تفاعل تداخلي بين أنواع الأسمدة وأنواع القمح على مستوى دلالة $\alpha = 0.05$.

هـ- استعمل اختبار شفييه لإيجاد أي أنواع القمح تختلف عن بعضها في التأثير على الإنتاج $\alpha = 0.05$.

و- استعمل اختبار شفييه لإيجاد أي أنواع الأسمدة تختلف عن بعضها في التأثير على الإنتاج. استعمل $\alpha = 0.05$.

ز- استعمل اختبار بونفيروني للإجابة عن السؤالين (هـ)، (و).

6-11 : استعملت أربعة أنواع من الأسمدة على ثلاثة أنواع من القمح زرعت في 12 قطعة أرض اختيرت عشوائيا مع افتراض نفس الخصوبة، فكان الإنتاج كما في الجدول التالي :

نوع السماد	نوع القمح		
	A	B	C
I	72	66	79
II	58	60	67
III	51	56	50
IV	61	60	68

أجب عن الأسئلة : أ ، ب ، جـ ، هـ ، و ، ز في تمرين 11-5.

7-11 : يعطي الجدول التالي الأوقات اللازمة للقطع حسب نوع الآلة (معاملات Treatment) وصلابة المادة (خلايا Blocks).

	صلابة المادة

الآلة	1	2	3	4	5	6
1	12	4	7	1	7	4
2	16	12	8	11	17	8
3	20	6	4	5	2	2
4	12	2	9	3	14	10

أ- ابن جدول تحليل التباين.

ب- هل يوجد فروق ذات دلالة على مستوى 0.05 بين الأوقات اللازمة للقطع تعزى لنوع الآلة المستعملة ؟

جـ- الفرع (ب) ولكن تعزى لصلابة المادة.

د- استعمل اختبار شفييه لإيجاد أي أنواع الآلات تختلف عن بعضها في الأوقات اللازمة للقطع.

هـ- استعمل اختبار t المتعدد لإيجاد أي أنواع صلابة المواد تختلف عن بعضها في الأوقات اللازمة للقطع.

8-11 : يمثل الجدول التالي نتيجة تحليل التباين لعرض السبلة لثلاثة أنواع من الأزهار حيث حجم العينة من كل نوع 50.

مصدر التغير	مجموع المربعات	درجات الحرية	معدل التغير	قيمة f
بين أنواع الأزهار	11.345			
الخطأ				
المجموع الكلي	28.307			

أ- أكمل الجدول واكتب الفرضية المناسبة ثم اختبرها.

ب- إذا علم أن المعدلات هي $\overline{Y}_1 = 3.428$, $\overline{Y}_3 = 2.974$, $\overline{Y}_2 = 2.770$.

اختبر أي المعدلات تختلف عن بعضها البعض.

9-11 : أكمل جدول تحليل التباين التالي :

مصدر التغير	مجموع المربعات	درجات الحرية	معدل التغير	قيمة f
بين المعاملات	12280			
الخطأ		12	423.17	
المجموع الكلي		14		

10-11 : أخذت عينات عشوائية من خمسة أنواع من الخيوط حجم كل عينة 3، وسجلت قوة الشدّ لكل خيط بالكغم فكانت النتيجة كما يلي :

أ	ب	جـ	د	هـ
20	17	22	18	13
22	19	20	16	12
18	12	21	17	14

1- ابن جدول تحليل التباين.

2- اختبر فرضية تساوي القوة للأنواع المختلفة للخيوط.

3- اختبر فيما إذا وجدت فروق في القوة بين كل نوعين من الخيوط.

11-11 يعطي الجدول التالي جدول تحليل التباين الثنائي ذي التفاعل :

مصدر التغير	مجموع المربعات	درجات الحرية	معدل المربعات	قيمة f
بين معدلات الصفوف	31.5706	3		
بين معدلات الأعمدة		2	7.302	
التفاعل بين الصفوف والأعمدة				

الخطأ	6.32			
المجموع الكلي	73.4438	47		

أ- أكمل الجدول.

ب- صغ جميع الفرضيات الصفرية التي يمكن اختبارها وأجر تلك الاختبارات.

الفصل الثاني عشر
السلاسل الزمنية

Time Series

1-12 : مقدمة

ينصب اهتمامنا في هذا الفصل على تحليل البيانات (المشاهدات المأخوذة) على فترة من الزمن. تسمى مثل هذه البيانات بالسلاسل الزمنية. وقد درس الاقتصاديون السلاسل الزمنية بشكل مكثف لأنها تدخل في كثير من المسائل الاقتصادية ذات الاهتمام. ولا تقتصر دراسة السلاسل الزمنية

علـى الأمـور الاقتصـادية ولكنهـا تشـغل مكانـة هامـة فـي العلـوم الطبيعيـة (كدراســة الطقــس والأمطــار والــزلازل والكــوارث الطبيعيــة) والصــناعة والتجارة ونمو السكان والتربية والرعاية الصحية وغيرها.

تعريف (1) : سلسلة المشاهدات المتتابعة عن ظاهرة معينة على مدى فترة زمنية تسمى سلسلة زمنية. تمثل البيانات في الجدول (1) والجدول (2) أمثلة على سلاسل زمنية.

الجدول (1) : حركة شحن البضائع في ميناء العقبة خلال السنوات 1976 وحتى 1989.

Year	عدد البواخر No. of Vessels
1976	1064
1977	944
1978	1197
1979	1238
1980	1466
1981	1744
1982	2599
1983	2454
1984	2329
1985	2671
1986	2677
1987	2555
1988	2583
1989	2446

الجدول (2) : النقل الجوي للشحن على الملكية الأردنية حسب الأشهر خلال السنوات 1985 وحتى 1987.

الشهر	الشحن بالطن		
	1987	1986	1985
كانون الثاني	3592.2	2944.7	2945
شباط	8322.7	3160.8	3299
آذار	4654.8	3839.0	3892
نيسان	4576.6	3488.2	3296
آيار	4320.4	3572.9	3778
حزيران	3342.1	3220.2	3429
تموز	3355.7	3366.4	3699
آب	3348.3	3122.6	3923
أيلول	3801.1	3663.8	3674
تشرين أول	4457.6	4345.9	3800
تشرين ثاني	4440.0	4197.4	3642
كانون الأول	4850.3	4378.6	3709
المجموع	48561.8	43300.7	43095

المصـدر: النشـرة الإحصـائية السـنوية/دائـرة الإحصـاءات العامـة/الأردن 1989 العدد 40.

إن القيم المشاهدة في السلسلة الزمنية مثل عدد الركاب أو كميات النقل أو ثمن سلعة معينة أو كمية المطر. عادة ما تكون نتائج لتأثيرات متعددة. إن اكتشاف هذه التأثيرات وقياس نتائجها تعتبران الأهداف الرئيسة في تحليل السلاسل الزمنية. ومع أن التأثيرات لا يمكن تحديدها بشكل يقيني مضبوط إلا أنه من الممكن إيجاد تقريب لها إذا ما كانت المشاهدات على فترة زمنية طويلة بشكل كاف. إن تحليل السلاسل الزمنية ذو أهمية في فهم السلوك في الماضي، وتقويم الانجازات في الحاضر وتخطيط العمليات في المستقبل إضافة إلى المقارنة بين السلاسل الزمنية المختلفة.

لقد جرت العادة باعتبار التغيرات في السلاسل الزمنية ناتجة من أربع مؤثرات محددة : الاتجاه العام، التغيرات الفصلية، التأثيرات الدورية والتغير العشوائي أو بسبب الخطأ.

وقد اصطلح على اعتبار كل واحد من هذه المؤثرات مركبة.

تعريف (2) : الاتجاه العام Secular Trend هو ذلك المؤشر أو تلك الخاصية للسلسلة الزمنية التي تمتد بشكل متناسق على مدى الفترة الزمنية تحت الدراسة.

إن مركبة الاتجاه العام هي المنحنى الرئيسي على المدى الطويل لنمو نشاط (ظاهرة) أو انكماشه، أي زيادته المطردة أو نقصانه. إنها تشير إلى وجود عوامل تستمر لمدة معقولة مثل عوامل نمو السكان، التطور التقني ، التغيرات في مستوى الأسعار، أو حالات مختلفة تخص بعض الصناعات أو المؤسسات.

تعريف (3) : التغيرات الفصلية Seasonal Variations : هي تغيرات تحدث في تتابع متسق على فترات زمنية محددة.

إن كلمة "فصلية" لا تعنى بالضرورة فصول السنة ولكنها تعني أي تغير له طبيعة دورية وتكون دوراته المتكررة ذات مدة زمنية قصيرة نسبيا. كثيرا ما تكون التغيرات الفصلية نتيجة عوامل فصلية مثل حالات الجو، العطل الرسمية، الأعياد، عمليات تجارية أو عادات المجتمع. مثال على ذلك أسعار السلع الزراعية القابلة للتلف حيث أن هذه الأسعار تكون عالية في بداية الموسم ثم تنخفض بشدة عند ذروة الإنتاج والقطف ثم ترتفع ثانية عند انتهاء الموسم.

وكمثال آخر، عدد الطلبة الذين يدرسون في المكتبة يزداد بشكل ملحوظ قبيل موعد الامتحان ثم ينقص بشدة بعد فترة الامتحانات.

تعريف (4) : التقلبات الدورية Cyclical Fluctuations

هي حركات طويلة الأمد تمثل الارتفاعات المتكررة والهبوطات المتكررة لنشاط أو ظاهرة ما.

إن التقلبات الدورية في السلاسل الزمنية للنشاطات التجارية عادة ما تسمى دورات اقتصادية. ومع أن التقلبات الدورية أقل عددا من التغيرات الفصلية إلا أن من أهم صفاتها أنها تحدث بانتظام. ويلاحظ أن معظم المؤسسات التجارية تعكس الفترات المتعاقبة من الركود والانتعاش الاقتصادي اللذين يمر بها الاقتصاد العام للدولة. وهناك دورات غير الدورات الاقتصادية مثل دورات كميات الإنتاج أو دورات الأسعار وغيرها.

تعريف (5) :

التغيرات العشوائية أو بسبب الأخطاء Random or erratic variations : وهي تلك التغيرات التي تحدث بشكل غير متنبأ.

تحدث التغيرات العشوائية لأسباب غير مرئية وغير متنبأة وربما تكون نتيجة الحروب أو الاضرابات أو الحرائق أو الزلازل أو الفيضانات أو الأحوال الجوية غير العادية أو الأحداث السياسية.

وباختصار فإن التغيرات العشوائية (أي غير المنتظمة) هي تلك التي لا تفسر عن طريق الاتجاه العام أو التغيرات الفصلية أو التقلبات الدورية.

2-12 : مركبـات السلاسـل الزمنيـة : Components of time series

كما ذكرنا في التعاريف (1) إلى (5) فإن مركبات السلسلة الزمنية هي :

(1) مركبة الاتجاه T.

(2) المركبة الفصلية S.

(3) مركبة الدورة C.

(4) المركبة غير المنتظمة I.

وهناك عدة نماذج تصف السلسلة الزمنية بدلالة مركباتها ومن النماذج الكلاسيكية النموذج الضربي أي حاصل الضرب ، اي أن :

$$Y = T \times X \times C \times I$$

حيث Y تمثل نتيجة السلسلة أو المشاهدة ، وهناك النموذج الجمعي، أي أن :

$$Y = T + S + C + I$$

ونماذج أخرى مثل $Y = TCS + I$

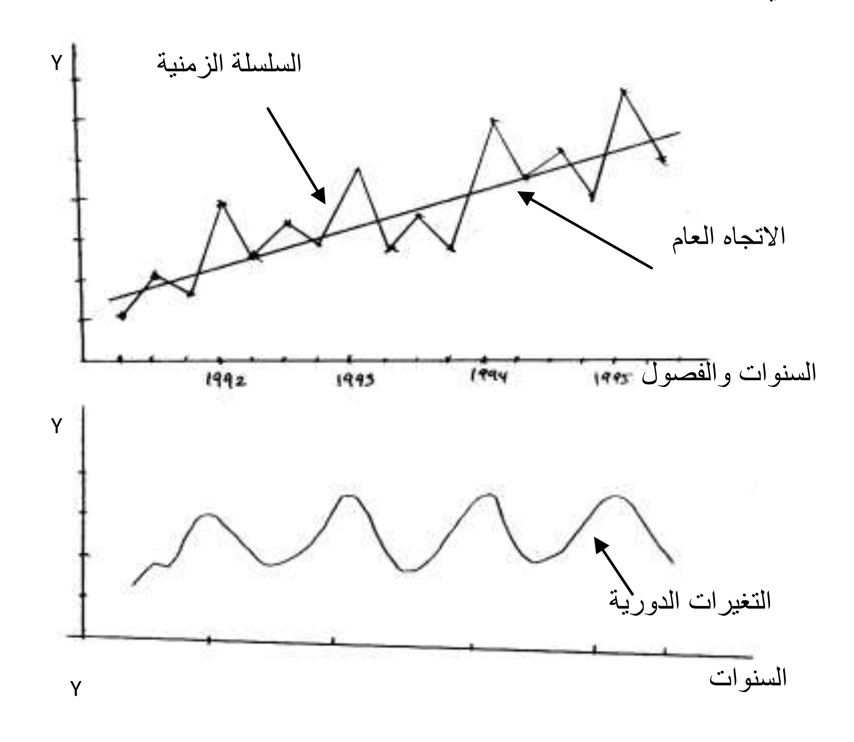
أو $Y = TC + SI$

وسنعتمد على تمثيل السلسلة الزمنية بالنموذج الضربي. ومع أن

كثيرا من المؤلفين والمحللين الاقتصاديين يعتمدون النموذج الجمعي إلا أن النموذج الضربي يمثل معظم السلاسل الاقتصادية.

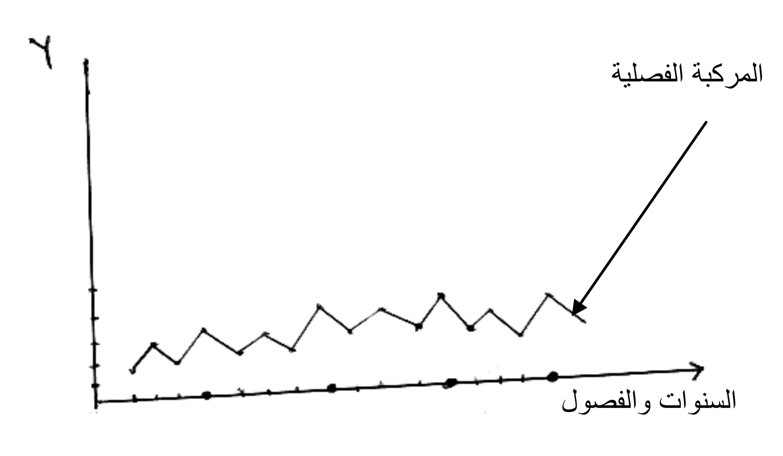
وبالتالي : "إن النموذج الضربي ليس فقط هو المعتمد كافتراض معياري أو تقليدي لتحليل السلاسل الزمنية بل هو الأكثر استعمالا في التطبيق العملي من جميع النماذج الأخرى مجتمعة" [1] .

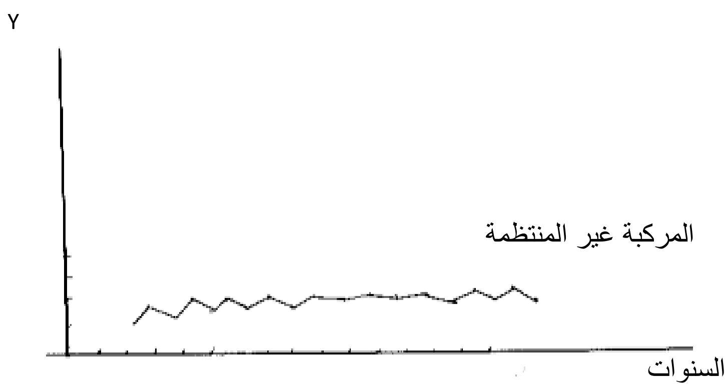
إن تحليل السلاسل الزمنية يعنى بإجراء عمليات إحصائية على البيانات لكي تفصل المركبات كل واحدة على حدة أي تجد قيمة كل واحدة منها.



---

[1] **Gupta, S.P. Statistical methods, Sultan Chard & Sons, 1981, New Delhi, India.**





3-12 : مركبة الاتجاه Trend

عندما تعطى سلسلة زمنية على مدة طويل (يعتمد تحديد المدى الطويل من طبيعة البيانات في السلسلة، فإذا كانت البيانات هي الصادرات أو الواردات مثلا فالمدى الطويل عدة سنوات أما إذا كانت البيانات عن تكاثر البكتيريا فالمدى الطويل ربما يكون ساعة واحدة حيث يمكن أخذ قراءات كثيرة في مدة ساعة، يكون هناك حاجة لمعرفة اتجاه هذه السلسلة هل هو متزايد أم متناقص.

إن قياس اتجاه السلسلة الزمنية مهم في ذاته ولذاته حيث بدراسة الاتجاه نحدد عامل النمو والذي بدوره يساعد في تنبوء مستقبل سلوك الظاهرة تحت الدراسة، ومن جهة ثانية فإن قياس الاتجاه يساعدنا في إلغاء دورة من السلسلة وبالتالي نتمكن من دراسة المركبات الأخرى.

هناك طرق عديدة لتحديد الاتجاه منها :

1- طريقة المعدل النصفي : Semi-Average Method

وتتلخص هذه الطريقة بتقسيم البيانات إلى قسمين متساويين وحساب معدل كل قسم ورصد كل معدل مقابل منتصف الفترة الزمنية التي حسب عليها ثم رسم خط مستقيم يصل بين هاتين النقطتين. فإذا كان عدد السنين (أو الأشهر مثلا) زوجيا كانت كل فترة تساوي نصف عدد السنين وإذا كان عدد السنين فرديا، حذفنا السنة المتوسطة، وقسمنا السنوات الباقية إلى قسمين متساويين.

مثال (1) :

أوجد اتجاه السلسلة الزمنية التي تمثل المستوردات في المملكة الأردنية للفترة الزمنية 1987-1972 جدول (3) بطريقة المعدل النصفي.

السنة	المستوردات
1972	95510.1
1973	108247.9
1974	156607.1
1975	234012.7
1976	339494.7
1977	454517.8
1978	458942.6
1979	585666.2
1980	715977.3
1981	1047505.2
1982	1142493.4
1983	1103310.4
1984	1071340.1
1985	1074445.4
1986	850199.2
1987	915554.7

الجدول (3)

المصدر: دائرة الجمارك

الحل : الفترة الزمنية تمثل 16 سنة، ولذلك نقسمها إلى قسمين كل منها 8 سنوات. نجد معدل القسم الأول (الفترة من 1972-1979) وهو :

$$\frac{95310.1+108247.9+.....+585666.2}{8}=\frac{2432799.1}{8}$$

$$= 304099.9$$

ونجد المعدل (الوسط الحسابي) للسنوات الثمان التالية فنجده :

$$\frac{715977.3+1047595.2+.....+915554.7}{8}=\frac{7920825.7}{8}$$

$$= 990103.2$$

نرصد على مستوى XY الديكارتي النقطتين : الأولى :

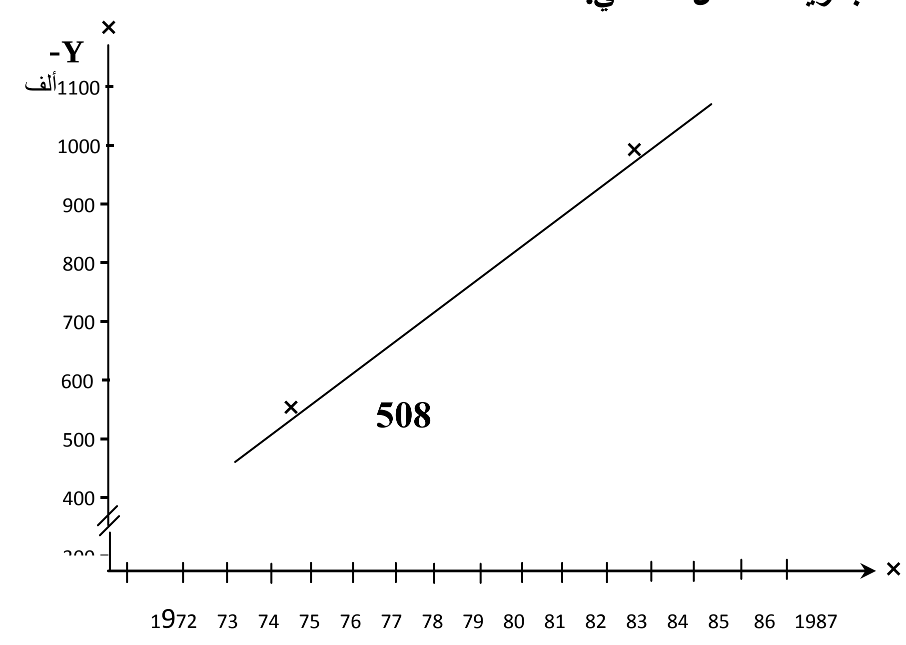
X = منتصف المسافة ما بين 1975 , 1976.

Y = 304099.9

الثانية : X = منتصف المسافة ما بين 1983، 1984.

Y = 990103.2.

ونصل بين النقطتين بخط مستقيم فيكون هو خط اتجاه السلسلة بطريقة المعدل النصفي.



2- طريقة المربعات الصغري : Least Sum of Squares

وهي الطريقة المستعملة في تعيين خط الانحدار البسيط وذلك بافتراض وجود علاقة خطية (أي تطبيق خط مستقيم على شكل الانتشار للسلسلة).

وبمراجعة معادلة خط الانحدار في الفصل العاشر ، نجد :

$$\hat{Y} = a + bX$$

حيث
$$b = \frac{\sum XY - n\bar{X}\bar{Y}}{\sum X^2 - n\bar{X}^2}$$

$$a = \bar{Y} - b\bar{X}$$

وبما أن السلسلة الزمنية عبارة عن مشاهدات أو بيانات Y مقابل الزمن، السنوات مثلا، الذي نعبر عنه بالرمز X ، فإنه يجب تعيين نقطة الأصل أو المركز، أي تعيين سنة محددة نضع X = 0 مقابلها وتعيين قيم X الأخرى بزيادة 1 أو طرح 1 حسب كون السنة تلي المركز أو قبله ، وهكذا.

مثال (2) :
لديك البيانات التالية :

جدول (4)

السنة	1988	89	90	91	92	93	94
الإنتاج	20	30	32	23	34	39	32

أ- وفق خط اتجاه مستقيم لهذه البيانات. وأرسمه على السلسلة الزمنية.

ب- كم تقدر إنتاج 1995.

الحل : نفرض المركز هو سنة 1988 فيكون مقابلها X = 0 ونكتب البيانات في جدول (5) ونجري الحسابات :

الجدول (5)

السنة	X	Y	X^2	XY
88	0	20	0	0
89	1	30	1	30
90	2	32	4	64
91	3	23	9	69
92	4	34	16	136
93	5	39	25	195
94	6	32	36	192
المجموع	21	210	91	686

$$\bar{X} = \frac{21}{7} = 3$$

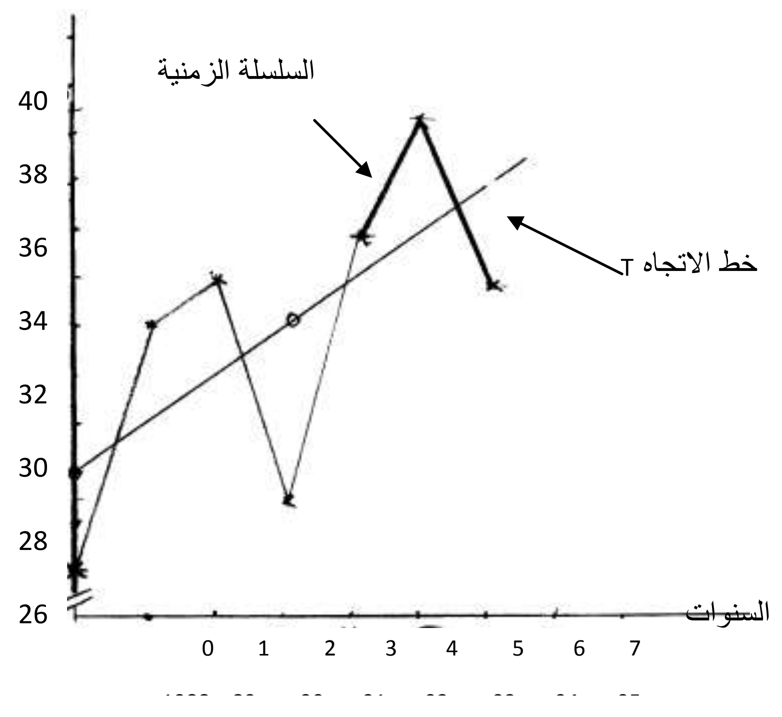
$$\bar{Y} = \frac{210}{7} = 30$$

$$b = \frac{686 - 7 \times 3 \times 30}{91 - 7 \times 3 \times 3}$$

$$= \frac{56}{28} = 2$$

$$a = 30 - 2 \times 3 = 24$$

إذاً $$T = \hat{Y} = 24 + 2X$$



لرسم المستقيم عوّض X = 1 تجد Y = 26

X = 4 تجد Y = 32

ارسم النقطتين (1. 26) ، (4. 32) وأوصل بينهما.

(ب) عام 1995 يقابلها X = 7

بافتراض أن الاتجاه يبقى متحققا فيكون تقدير إنتاج 1995 هو :

$T_{95} = 24 + 2 \times 7 = 38$

مثال (3) :

في المثال (2) خذ المركز سنة 1991 وحل المثال :

الحل : نرتب الحل كما في الجدول (6).

الجدول (6)

السنة	X	Y	X^2	XY
88	-3	20	9	-60
89	-2	30	4	-60
90	-1	32	1	-32
91	0	23	0	0
92	1	34	1	34
93	2	39	4	78
94	3	32	9	96
المجموع	0	210	28	56

$$\mathbf{b = \frac{56 - 7 \times 0 \times 30}{28 - 0} = 2}$$

$$\mathbf{a = 30 - 2 \times 0 = 30}$$

إذا خط الاتجاه يكون :

$$T = 30 + 2X$$

ويمكن رسم هذا الخط ولكن باعتبار المركز X = 0 يقابل عام 1991. وهو نفس الشكل السابق ولكن إحداثي Y يكون على 1991.

(ب) لتقدير إنتاج 1995،نضع X = 4 التي تقابل عام 95 ويكون $T_{95} = 30 + 2 \times 4$ = 38 = 4 كما في المثال (2).

ومن هذين المثالين نلاحظ أن ميل خط الانحدار (خط الاتجاه) يبقى نفسه حتى لو اختلف المركز من عام لآخر وكذلك التنبؤ يبقى نفسه.

نلاحظ أن الاختلاف في معادلتي الانحدار في المثالين هو قيمة Y عندما X = 0 أما الميل (قيمة b) فتبقى نفسها ويمكننا إيجاد معادلة الانحدار (أو الاتجاه) عند تغيير المركز مباشرة وذلك من الملاحظة (i) :

(i) تغيير مركز الاتجاه : Shifting the Trend Origin

كما ذكرنا سابقا، عندما نريد إيجاد معادلة الاتجاه فلابد من تحديد المركز (السنة التي يقابلها X = 0) وعادة ما نأخذ السنة المتوسطة أو السنة الأولى ليقابلها X = 0.

فإذا كانت معادلة الاتجاه الخطي :

$$T = a + bX$$

وسحبنا المركز k من الوحدات فإن المعادلة تصبح $T = a + b\,(X + k)$

مع ملاحظة أن k موجبة إذا غيرنا المركز إلى سنة بعد (تلي) سنة المركز الأصلية وتكون سالبة إذا غيّرنا المركز إلى سنة تسبق سنة المركز الأصلية ففي المثالين (2)، (3) :

عندما كان المركز عام 1988 كانت المعادلة : $T = 24 + 2X$

وعندما غيرنا المركز إلى 1991 أي $k = 3$ تصبح المعادلة :

$$Y = 24 + 2\,(X + 3)$$

$$= 30 + 2X$$

وهناك أمر آخر هام في دراسة الاتجاه وهو الملاحظة (ii) :

(ii) تحويل خط الاتجاه من قيم سنوية إلى قيم شهرية

Conversion of Annual Trend Values to Monthly Values

يمكننا تحويل خط الاتجاه السنوي إلى خط اتجاه شهري وذلك بقسمة a على 12 وقسمة b على $(12)^2$، وبذلك فالمعادلة :

$$T_1 = a + b\,X$$

X وحدة السنة تصبح :

$$T_2 = \frac{a}{12} + \frac{b}{144}X$$

X وحدة الشهر.

مثال (4) : في مثال (2) ، أوجد خط الانحدار (الاتجاه) إلى أساس شهري.

الحل :

T = 24 + 2X

X : وحدة السنة.

تصبح $$T_2 = \frac{24}{12} + \frac{2}{144}X$$

أي $$T_2 = 2 + \frac{1}{72}X$$

X وحدة الشهر.

3- طريقة المعدلات المتحركة Methods of Moving Averages

عندما نريد إيجاد الاتجاه بطريقة المعدلات المتحركة فإننا نحسب معدل المشاهدات لعدد من السنين (أو الفصول أو الأشهر) ونعتمد هذا المعدل قيمة للاتجاه مقابلا لوحدة الزمن الواقعة في منتصف الفترة التي حسب المعدل على أساسها. إن التعديل (أخذ المعدل) يعطي منحنى أكثر تمهيدا (أملس) وبالتالي يقلل من تأثير التقلبات التي تسحب قيم المشاهدات السنوية عن خط الاتجاه العام.

وعادة ما نأخذ المعدل المتحرك بطول ثلاث سنوات أو خمس سنوات او 8 سنوات وذلك اعتمادا على طول الدورة التي نرى أن السلسلة الزمنية واقعة تحت تأثيرها.

وقبل حساب الاتجاه بطريقة المعدلات المتحركة نلفت الانتباه أن طريقة المعدلات المتحركة يمكن استعمالها في حساب المركبة الفصلية ومركبة الدورة ومركبة الخطأ كما سنرى في البنود القادمة.
ونشرح هذه الطريقة كما في الأمثلة.

أولا : المعدلات المتحركة بطول فردي :

مثال(5) : أوجد المعدلات المتحركة بطول 3 للسلسلة الزمنية المعطاة في الجدول (7)

السنة	1980	1981	1982	1983	1984	1985	1986
المشاهدة	15	12	9	18	15	24	27

جدول (7)

الحل : نرتب الحل كما في الجدول (8).

الجدول (8)

السنة العمود 1	المشاهدات العمود 2	مجموع المشاهدات بطول 3	معدلات المشاهدات بطول 3
1980	15	-	-
81	12	36	12
82	9	39	13
83	18	42	14
84	15	57	19
85	24	66	21
86	27	-	-

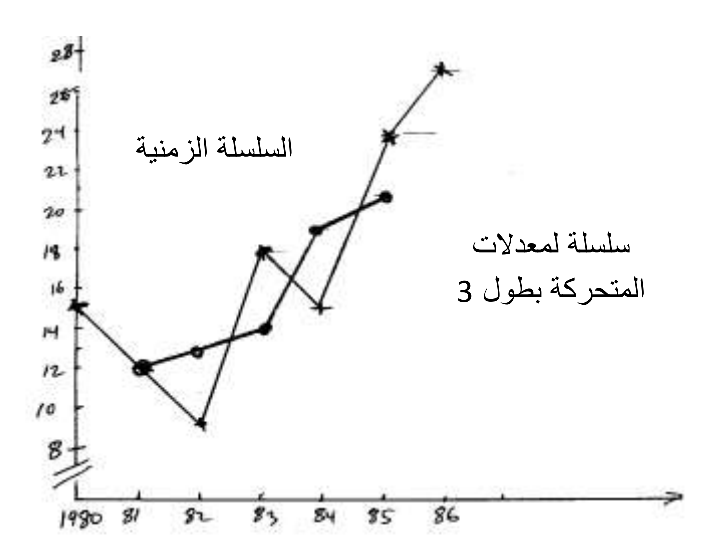
في مثال (5) أخذنا الطول 3 سنوات.

نملأ العمود (3) وهو المجموع بطول 3 سنوات وذلك بأخذ مجموع المشاهدات للسنوات الثلاث الأولى ثم نترك السنة الأولى ونأخذ السنة الرابعة ونجد مجموع المشاهدات لهذه السنوات الثلاث وهكذا.

نملأ عمود (4) بأن نأخذ الوسط الحسابي أي بقسمة كل رقم في عمود (3) على 3.

إن العمود (4) يمثل المعدلات المتحركة بطول 3 سنوات وعندما نرصد هذه المعدلات على نفس شكل الانتشار للسلسلة الزمنية نلاحظ كيف تم تمهيدها فأصبحت ملساء نوعا ما وتعطينا هذه المعدلات اتجاه السلسلة.

السنة



ثانيا : المعدلات المتحركة بطول زوجي

مثال (6) : أوجد المعدلات المتحركة المركزية بطول 4 للسلسلة الزمنية المعطاة في الجدول (9) (الأعمدة 1، 2، 3).

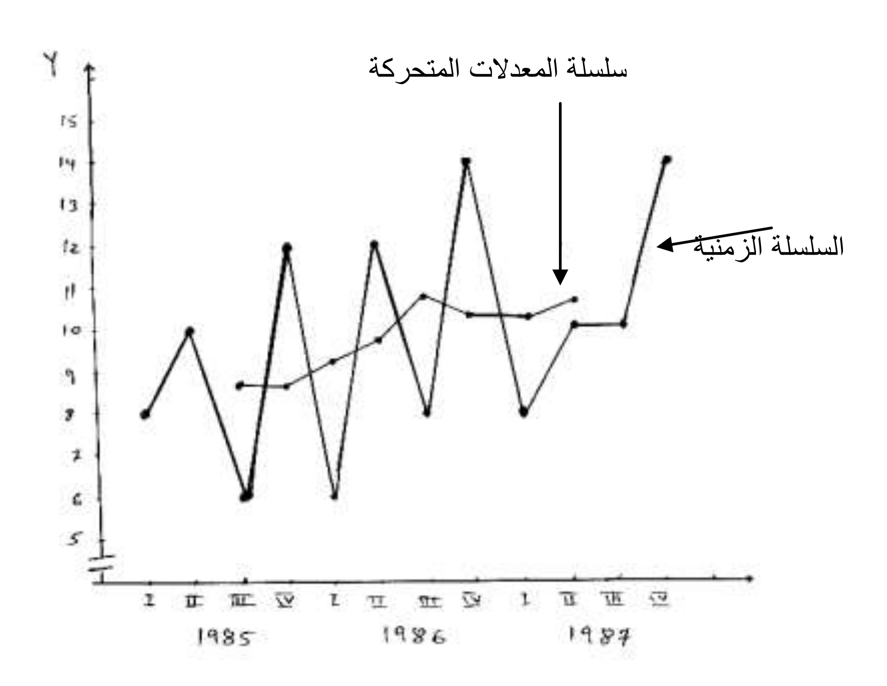
الجدول (9)

السنة العمود 1	الفصل العمود 2	المشاهدة العمود 3	المجموع بطول 4	المعدل بطول 4	المعدل المركزي بطول 4
1985	I	8			
	II	10	36	9	8.75
	III	6	34	8.5	8.75
	IV	12			
1986	I	6			9.25

	II	12	36	9	9.75
	III	8	38	9.5	
	IV	14	40	10	10.75
			42	10.5	10.25
1987	I	8	40	10	10.25
	II	10	42	10.5	10.5
	III	10	42	10.5	
	IV	14			

لاحظ أننا نبني العمود (4) بأخذ مجموع المشاهدات الأربع الأولى $(Y_1 + Y_2 + Y_3 + Y_4)$ ثم نترك المشاهدة الأولى ونأخذ المشاهدة الخامسة ونجد مجموع هذه المشاهدات الأربع أي نجد $Y_2 + Y_3 + Y_4 + Y_5$ وهكذا.

نملأ العمود (5) بأخذ معدل كل مجموع في العمود (4) أي بالقسمة على 4 ونلاحظ أن المدخلات في العمودين (4)، (5)، تقع بين الفصول. نملأ العمود (6) بأخذ الوسط الحسابي لكل معدلين وتكون المدخلات في هذا العمود مقابلة للفصول. تمثل القيم في العمود (6) المعدل المركزي بطول 4 وبرصد هذه القيم على نفس شكل الانتشار للسلسلة الزمنية يظهر اتجاه تلك السلسلة كما في الشكل التالي.



إن المعدلات المتحركة في مثال (5) والمعدلات المتحركة المركزية في مثال (6) تعطينا اتجاه السلسلة في كل منهما ويمكننا رصد المعدلات المتحركة على شكل الانتشار للسلسلة لنلاحظ اتجاه تلك السلسلة كما وضحنا في المثالين (5)، (6) وهناك فائدة أخرى للمعدلات المتحركة وهي استخلاص مركبة التذبذب في السلسلة وتعرف هذه المركبة على أنها "المشاهدة في السلسلة ـ المعدل المتحرك المقابل لها" ، ويكون هناك حالتان :

(1) المعدل المتحرك بطول فردي.

(2) المعدل المتحرك المركزي بطول زوجي.

في مثال (5) تكون مركبة التذبذب :

الجدول (10)

السنة	1980	1981	1982	1983	1984	1985	1986

المشاهدة	15	12	9	18	15	24	27
المعدل المتحرك		12	13	14	19	21	
مركبة التذبذب		0	4-	4	-4	3	

وفي مثال (6) تكون مركبة التذبذب كما يلي :

الجدول (11)

السنة	1985				1986				1987			
	I	II	III	IV	I	II	III	IV	I	II	III	IV
المشاهدات	8	19	6	12	6	12	8	14	8	10	10	14
المعدل المركزي بطول 4	-	-	8.75	8.75	9.25	9.75	10.25	10.25	10.25	10.25	-	-
مركبة التذبذب	-	-	-2.75	3.25	3.25 -	2.25	2.25 -	3.75	2.25 -	0.25 -	-	-

4-12 : تقـدير المركبـة الفصـلية : Measurement of Seasonal Variations

إن معظم الظواهر في الاقتصاد والتجارة يظهر فيها تقلبات فصلية، وعندما تكون البيانات مسجلة سنويا فلا يوجد تغيرات فصلية، ولكن البيانات الشهرية أو الفصلية غالبا ما تظهر تقلبات فصلية وبالتالي تظهر الحاجة إلى إيجاد نمط معدل التغير الفصلي.

وهناك طرق متعددة لقياس أو تقدير المركبة الفصلية ندرس أهمها :

(1) النسبة إلى المعدل المتحرك : Ratio to Moving Average

نفترض في هذه الطريقة أن نموذج السلسلة الزمنية هو النموذج الضربي $Y = T \times S \times C \times I$ ونفترض أن المركبة الفصلية S لها دورة باثنى عشر شهرا(أو بأربعة فصول) وأن شكل التغير يبقى نفسه لكل سنة. ونفترض أيضا أن التغيرات العشوائية (I) مستقلة عن بعضها البعض للفترات المختلفة.

عندما نفترض هذه الافتراضات ونحسب المعدلات المتحركة بطول 12 شهرا (أو 4 فصول) فإن المعدل المتحرك يكون عبارة عن $T \times C$ وبقسمة البيانات الأصلية على المعدل المتحرك نحصل على $S \times I$ أي أن :

$$\frac{\text{المشاهدات الأصلية}}{\text{المعدل المتحرك}} = \frac{T \times S \times C \times I}{T \times C} \text{ ويساوي } S \times I$$

وبالحصول على $S \times I$ نرغب في إزالة ما نستطيع من تأثير (I) على $S \times I$ ويكون ذلك بأخذ معدل $S \times I$ ومن ثم جعل هذا المعدل نسبة مئوية وهي التي تمثل المركبة الفصلية للشهر المعين الذي حسبت له. وملخص هذا في الخطوات التالية :

(i) نجد المعدل المتحرك المركزي بطول مناسب (12 إذا كانت البيانات شهرية) أو 4 إذا كانت فصلية.

(ii) نقسم البيانات الأساسية على المعدل المتحرك المركزي ونضرب الناتج في 100 فنحصل على حاصل ضرب مركبتي الفصل والخطأ $S \times I$.

(iii) نعزل مركبة الفصل عن مركبة الخطأ وذلك بالتعديل على السنوات المختلفة.

(iv) إذا كان مجموع المركبات الناتجة في الخطوة الثالثة أعلاه هو B تكون مركبة الفصل تساوي :

$$\frac{\text{نتيجة الفصل في الخطوة الثالثة}}{B} \times \text{عدد الفصول} \times 100$$

مثال (7) : يبين الجدول 12 (العمودان 1 ، 2) النقل الجوي للركاب على الملكية الأردنية حسب الفصول الأربعة خلال السنوات 1987 وحتى 1991.

المصدر : الملكية الأردنية / النشرة الإحصائية السنوية / دائرة الإحصاءات العامة/ الأردن 1991 العدد 42.

الجدول (12)

السنة والفصول الأربعة (1)		المشاهدات (عدد الركاب) (2)	المجموع لأربعة فصول (3)	المعدل المتحرك لأربعة فصول (4)	المعدل المتحرك المركزي لأربعة فصول (5)	العمود(2) ـــــ × 100 العمود(5) = S × I
	1	222734			-	
1987	2	256571	1119654	279913.5	-	
	3	398495	1165121	291280.3	285596.9	139.5
	4	241854	1188295	297073.8	294177.1	82.2
	1	268201	1204084	301021	299047.4	89.7
1988	2	279745	1225934	306483.5	303752.3	92.1
	3	414284	1232091	308022.8	307253.2	134.8
	4	263704	1240291	310072.8	309047.8	85.3
	1	274358	1230354	307588.5	308830.7	88.8
1989	2	287945	1204005	301001.3	304294.9	94.6
	3	404347	1175391	2938478	297424.6	135.9
	4	237355	1163816	290954	292400.9	81.2
	1	245744	1061086	265171.5	278112.8	88.4
1990	2	276370	963915	240978.8	253125.2	109.2
	3	301617	785619	196404.8	218691.8	137.9
	4	140184			183993.1	76.2

	1	67448	686325	171581.3	174103.9	38.7
1991	2	177076	706506	176626.5	188049.8	94.2
	3	321798	797892	199473	-	-
	4	231570			-	-
						1569.4

الجدول (13)

السنة	الفصل (الربع) الأول	الربع الثاني	الربع الثالث	الربع الرابع
1987	ــــ	ــــ	139.5	82.2
1988	89.7	92.1	134.8	85.3
1989	88.8	94.6	135.9	81.2
1990	88.4	109.2	137.9	76.2
1991	38.7	94.2	ــــ	ــــ
المعدل	76.4	97.5	137.0	81.2
المجموع				392.1

يبين الجدول(12) الحسابات في الخطوتين (i) ، (ii) للمسلسلة المعرفة في العمودين (1) ، (2).

ولحساب الخطوتين (iii) ، (iv)، نرتب البيانات في العمود الأخير من جدول (12) ونضعها في جدول (13) ونجد معدل كل فصل.

مجموع المعدلات B = 392.1

إذاً مركبة الفصل الأول تساوي : $\frac{76.4}{392.1} \times 400 = 77.94$

مركبة الفصل الثاني : $\frac{97.5}{392.1} \times 400 = 99.46$

مركبة الفصل الثالث : $\frac{137.0}{392.1} \times 400 = 139.76$

مركبة الفصل الرابع : $\frac{81.2}{392.1} \times 400 = 82.84$

إن تفسير مركبة الفصل بسيط حيث نقول (في مثالنا) إن عدد المسافرين على الملكية الأردنية في الربع الأول (الأشهر الثلاثة الأولى من السنة) يساوي 77.94 بالمائة من معدل عدد المسافرين في الربع الواحد، وهكذا.

البيانات اللافصلية Deseasonalized Data

البيانات اللافصلية هي البيانات التي تظهر لنا كيف يمكن أن تكون المشاهدات لو أنه لم يكن هناك تأثيرات فصلية، وللحصول على مثل هذه البيانات علينا إزالة تأثير المركبة الفصلية منها.

وطريقة إزالة تأثير المركبة الفصلية تكون بقسمة المشاهدة الفصلية لفصل معين على مركبة ذلك الفصل ثم الضرب بمائة. بالطبع إذا كانت البيانات شهرية واستخرجنا المركبة الشهرية (المركبة الفصلية لكل شهر) نجعل البيانات لافصلية بقسمة المشاهدة على المركبة الشهرية والضرب بمائة. فمثلا مشاهدات عام 1987 غير الفصلية (اللافصلية) هي :

الفصل الأول (الربع) الأول : $\frac{222734}{77.94} \times 100 = 285776$

الفصل الثاني : $\frac{256571}{97.5} \times 100 = 263150$

وهكذا.

أما مشاهدات عام 1991 اللافصلية فهي كما في الجدول (14).

الجدول (14)

(1) الربع	(2) المشاهدة	(3) مركبة الفصل	القيمة اللافصلية = (العمود (2) ÷ العمود (3)) × 100
الأول	67448	77.94	86538
الثاني	177076	99.46	178037
الثالث	321798	139.76	230250
الرابع	231570	82.84	279539

2- النسبة إلى الاتجاه Ratio to Trend

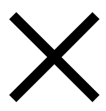
وتعتبر هذه الطريقة الثانية في الأهمية بعد طريقة النسبة إلى المعدل المتحرك، وتفترض الافتراضات نفسها وتحسب بطريقة مشابهة كما يلي :

(i) نجد معادلة الاتجاه بطريقة المربعات الصغرى، $T = \hat{Y} = a + bX$.

ونجد القيمة التقديرية (تنبوء قيمة) لكل فصل (ربع) أو شهر في السلسلة الزمنية أي نجد قيمة الاتجاه لكل فصل أو شهر.

(ii) نقسم البيانات الأساسية على قيمة الاتجاه ونضرب الناتج في 100 فنحصل على :

$$\frac{\mathbf{S \times T \times C \times I}}{\mathbf{T}} = \mathbf{S \times C \times I}$$



(iii) نعزل مركبة الفصل S عن I C وذلك بالتعديل (أخذ الوسط الحسابي) على السنوات المختلفة.

(iv) إذا كان مجموع المركبات الناتجة في الخطوة الثالثة أعلاه هو B تكون مركبة الفصل تساوي :

$$\frac{\text{نتيجة الفصل في الخطوة الثالثة}}{B} \times \text{عدد الفصول} \times 100$$

مثال (8) : بطريقة النسبة إلى الاتجاه احسب مركبة الفصل للسلسلة الزمنية في الجدول (15) العمودين (1)، (2) التي تمثل عدد زوار متحف الآثار الأردني في مدينة عمان / المصدر في مثال (7) حسب الفصول خلال الأعوام 1988-1991.

الجدول (15)

(1) العام والفصول الأربعة		(2) عدد الزوار Y	(3) X	(4) X^2	(5) XY	(6) T =	$\frac{\text{العمود}(2)}{\text{العمود}(6)} \times 100$
	1	5652	-7	49	-39564	7986.52	70.77
1988	2	5282	-6	36	-31692	7708.51	68.52
	3	4538	-5	25	-22690	7430.50	61.07
	4	6354	-4	16	-25416	7152.49	88.84
	1	11141	-3	9	-33423	6874.48	162.06
1989	2	6938	-2	4	-13876	6596.65	105.17
	3	4621	-1	1	-4621	6318.46	73.13
	4	12249	0	0	0	6040.45	202.78
	1	11969	1	1	11969	5762.44	207.71
1990	2	9996	2	4	19992	5484.43	182.26
	3	1976	3	9	5928	5206.42	37.95
	4	4109	4	16	16436	4928.41	37؟83
	1	841	5	25	4205	4650.4	18.08
1991	2	1051	6	36	6306	4372.39	24.04
	3	2515	7	49	17605	4094.38	61.43
	4	5191	8	64	41528	3816.37	136.02

		94423	8	344	-47313		

الحل: نضع الحل بإكمال الجدول (15) بادئين من العمود (3) فنحسب b ، a لنجد معادلة الانحدار (الاتجاه).

$$b = \frac{-47313 - 16 \times \frac{8}{16} \times \frac{94423}{16}}{344 - 16 \times \left(\frac{1}{2}\right)^2}$$

$$= \frac{-94524.5}{340} = -278.01$$

$$a = 5901.44 - (-278.01) \times \frac{1}{2}$$

$$= 6040.45$$

$$T = \hat{Y} = 6040.45 - 278.01X$$

حيث X = ربع سنة.

و 0 = X تقابل الربع الرابع لسنة 1989.

ومن معادلة الانحدار (الاتجاه) نجد قيمة $\hat{Y}$ لكل فصل بتعويض قيمة X المقابلة لذلك الفصل في معادلة الانحدار، فمثلا للفصل الأول لعام 1988 نجد :

$$T = \hat{Y} = 6040.45 - 278.01 \times -7 = 7986.52$$

وللفصل الثاني :

$$T = 6040.45 - 278.01 \times -6 = 7708.51$$

ثم نجد العمود (7) بقسمة العمود (2) على العمود (6) ثم الضرب بمائة. نضع مدخلات العمود (7) حسب الفصول والسنوات كما في الجدول (16) ثم نجد معدل كل فصل ونجمع المعدلات وهو = B 395.81 والآن نجد المركبة الفصلية كما يلي :

معدل الفصل الأول

مركبة الفصل الأول = ________ × 4 × 100 = 115.87

395.81

ونحسب مركبة كل فصل بنفس الطريقة ونضع النتيجة في الجدول (17).

الجدول (16)

السنة	الفصل (الربع) الأول	الثاني	الثالث	الرابع	
1988	70.77	68.51	61.07	88.84	
1989	162.06	105.17	73.13	202.78	
1990	207.71	182.26	37.95	83.37	
المجموع	458.62	379.99	233.58	511.01	
المعدل	114.66	95.00	58.40	127.73	395.81

الجدول (17)

المركبة الفعلية				المجموع
الربع الأول	الربع الثاني	الربع الثالث	الربع الرابع	
115.87	96.01	59.02	129.08	399.98

المفروض أن يكون المجموع 400 ولكن اختلف المجموع عن ذلك بسبب التقريب.

بعد إيجاد المركبة الفصلية نستطيع أن نعزل قيم هذه المركبة ونحصل على القيم اللافصلية للمشاهدات والقيم اللافصلية هي عبارة عن حاصل ضرب مركبات الاتجاه، الدورة، والأخطاء، أي $T \times C \times I$.

$$\text{والقيمة اللافصلية} = \frac{\text{المشاهدة للفصل المعين}}{\text{المركبة الفصلية له}} \times 100$$

مثال (9) : اكتب القيم اللافصلية لعدد الزوار عام 1988 في مثال (8).

الحل : كما في الجدول (18).

العمود (4) = (العمود (2) / العمود (3)) × 100

الجدول (18)

الفصل (1)	المشاهدة (2)	المركبة الفصلية (3)	القيمة اللافصلية (4)
1	5651	115.87	4879
2	5282	96.01	5502
3	4538	59.02	7689
4	6354	129.08	4923

3- طريقة المعدلات البسيطة Method of Simple Averages :

وهذه الطريقة أسهل طرق إيجاد المركبة الفصلية ولكنها غير دقيقة وهي الأقل استعمالا ونشرحها هنا لاستكمال الموضوع وتتلخص بإيجاد معدل المشاهدات لكل فصل (ربع سنة أو شهر أو غيره)، ثم جمع هذه المعدلات فإذا كان مجموعها B فإن :

مركبة الفصل = (معدل ذلك الفصل / B) × عدد الفصول × 100

مثال (10) : يظهر الجدول (19) معدل درجات الحرارة المئوية م° الشهري في مطار عمان خلال الفترة 1989 - 1985.

أوجد المركبة الفصلية حسب الأشهر بطريقة المعدلات البسيطة.

الحل : كما في الجدول (19)، نجد الوسط الحسابي لمعدل درجات الحرارة لكل شهر، ثم نجمع هذه الأوساط فنجد :

$$B = 8.4 + 9.2 + 11.5 + ... + 9.7 = 210.1$$

المركبة الفصلية لشهر كانون الثاني (يناير) يساوي : $\frac{12 \times 8.4}{210.1} \times 100 = 47.98$

وبنفس الطريقة نحسب المركبة الفصلية للأشهر جميعها كما في الجدول (20).

معدل درجة الحرارة "م°" الشهري في مطار عمان خلال الفترة 1985-1989.

الجدول (19)

السنة	كانون الثاني	شباط	أذار	نيسان	أيار	حزيران	تموز	آب	أيلول	تشرين الأول	تشرين الثاني	كانون الأول
1985	10.4	07.7	12.0	15.6	21.2	23.6	24.9	28.1	23.9	18.0	16.4	10.1
1986	08.6	10.5	13.4	18.6	18.0	23.5	25.6	25.6	25.8	20.1	11.6	08.4
1987	09.0	11.2	09.1	14.9	21.3	23.7	26.5	26.4	24.3	18.7	14.9	10.3
1988	08.2	08.8	11.1	16.5	22.1	24.3	26.8	25.7	24.4	19.2	12.2	09.7
1989	05.6	07.9	11.7	20.3	21.7	23.1	25.4	25.5	23.6	19.6	15.1	09.9
المجموع	41.8	46.1	57.3	85.9	104.3	118.2	129.2	131.3	122.0	95.6	70.2	48.4

المعدل	8.4	9.2	11.5	17.2	20.9	23.6	25.8	26.3	24.4	19.1	14.0	9.7

المصدر: دائرة الأرصاد الجوية /، النشرة الإحصائية السنوية / دائرة الإحصاءات العامة. الأردن 1989 العدد 40.

المركبة الفصلية حسب الأشهر

الجدول (20)

الشهر	كانون الثاني	شباط	أذار	نيسان	أيار	حزيران	تموز	آب	أيلول	تشرين الأول	تشرين الثاني	كانون الأول
المركبة	47.98	52.55	65.68	98.34	119.37	134.79	147.36	150.21	139.36	109.09	79.96	55.40

وتحسب القيم اللافصلية كالسابق :

$$\text{القيمة اللافصلية لشهر تموز (يوليو)} = \frac{\text{معدل شهر تموز} \times 100}{\text{المركبة الفصلية لشهر تموز}}$$

$$= \frac{25.8 \times 100}{147.36} = 17.51$$

وهي معدل درجة الحرارة لشهر تموز فيما لو عزل منه تأثير المركبة الفصلية لذلك الشهر.

5-12 : قيـاس التغيـرات الدوريـة Measurement of Cyclical Variation:

تعتبر الدورات الاقتصادية من أهم التقلبات في البيانات التجارية، وقد اهتم الباحثون بدراسة هذه التقلبات. وبالرغم من أهمية الدورات الاقتصادية فإنها من أصعب التقلبات تقديرا وقياسا. ولاختلاف هذه التقلبات الاقتصادية في توقيتها وحدتها ونظام حدوثها فمن الصعب إيجاد قياسات دقيقة Indexes لها كما هو الحال في قياس الاتجاه والمركبة الفصلية.

ومن أهم طرق قياس التقلبات الدورية طريقة البواقي.

طريقة البواقيMethod Residual

هذه الطريقة هي الأكثر استعمالا من بين طرق تقدير المركبة الدورية وتتلخص في عزل مركبتي الاتجاه والفصل أي :

$$\frac{T\times S\times C\times I}{T\times S}=C\times I$$

فنحصل على مركبتي الدورة والخطأ.

بعد ذلك نقوم بتمهيد البيانات للحصول على التقلبات (الحركات) الدورية.

6-12 : قياس مركبة الخطأ (التغيرات غير المنتظمة)

Measurement of Irregular Variations

إن مركبة الخطأ في السلسلة الزمنية تمثل ما يتبقى من التغيرات بعد عزل مركبات الاتجاه والفصل والدورة أي :

$$\frac{T\times S\times C\times I}{T\times S\times C}=I$$

وفي الأحوال العملية فإن الدورة نفسـها تكون متشـابكة مـع التغيـرات غير المنتظمة لدرجة أنه يصعب (عزل) فصل الواحدة عن الأخرى. وبالتـالي عنـد تحليـل السلسـلة الزمنيـة علـى مركباتهـا فإننـا نقـيس أو نقـدر كـلا مـن مركبتي الاتجاه والفصل بشكل مباشر ونترك مركبتي الدورة والخطأ معا بعد عزل المركبتين الأخريين.

7-12 : تنبؤ قيم السلسلة الزمنية :

Forecasting of Values of a Time Series

لقد اعتمدنا النموذج الضربي لتمثيل السلسلة الزمنية أي $Y = T \times S \times C \times I$ فمثلا إذا كان لدينا سلسلة زمنية عند الزمن t فيها $C = 86$, $S = 117$, $T = 120000$, $I = 71$ فإن قيمة السلسلة عند الزمن t تكون :

$$\mathbf{Y = 120000 \times \frac{117}{100} \times \frac{86}{100} \times \frac{71}{100}}$$
$$\mathbf{= 85728.24}$$

أي أننا نسترجع قيمة السلسلة عند زمن معين إذا عرفنا قيم مركباتها عند ذلك الزمن، مع ملاحظة أن المركبات I , C , S هي نسب مئوية.

إذاً نستطيع أن نتنبأ قيمة السلسلة إذا افترضنا أن مركبات السلسلة تبقى متحققة أي تكون الظروف المستقبلية تشبه الظروف التي حسبنا فيها المركبات.

فإذا عرفت المركبة الفصلية وقيمة اتجاه الفصل فإن :

$$\text{القيمة المتوقعة للفصل} = \text{قيمة اتجاه ذلك الفصل} \times \frac{\text{المركبة الفصلية له}}{100}$$

مثال (11) : إذا كانت معادلة خط الانحدار (الاتجاه) لسلسلة زمنية من عام 1988 وحتى عام 1997 هي :

$$\mathbf{\hat{Y} = 12000 + 200X}$$

حيث X وحدة الفصل، ونقطة المركز هي $X = 0$ تمثل الفصل الأول لعام 1988.

أوجد القيمة المتوقعة للفصل الثاني لعام 1998 إذا كانت مركبة الفصل الثاني 97.8 وبافتراض تحقق الظروف نفسها.

الحل : الفصل الثاني لعام 1998 يقابله :

$X = 10 \times 4 - 1 + 2 = 41$

وذلك لوجود 10 سنوات في كل منها 4 فصول ويوجد فصلان في عام 1998 ونطرح 1 لأننا بدأنا المركز من الفصل الأول $X = 0$.

إذا قيمة الاتجاه للفصل الثاني لعام 1998 هو :

$$\mathbf{T = \hat{Y} = 12000 + 200 \times 41 = 20200}$$

إذا القيمة المتوقعة تساوي :

$$\mathbf{\frac{20200 \times 97.8}{100} = 19755.6}$$

مثال (12) : في مثال (8) أوجد تقديرا لعدد الزوار لمتحف الآثار الأردني في مدينة عمان (أ) للفصل الثالث من عام 1993 (ب) للفصل الرابع من نفس العام.

الحل : من مثال (8) فإن معادلة الاتجاه هي :

$$\mathbf{T = \hat{Y} = 6040.45 - 278.01X}$$

حيث X = ربع سنة و X = 0 تقابل الربع الرابع لعام 1989.

إذاً الفصل الثالث لعام 1993 يقابله

$X = 3 \times 4 + 3 = 15$

إذا :

$T = 6040.45 - 278.01 \times 15$

$= 1870.3$

وبما أن مركبة الفصل الثالث هي 59.02

إذا القيمة المتوقعة تساوي :

$$\mathbf{\frac{1870.3 \times 59.02}{100} \cong 1104}$$

عدد الزائرين المتوقع.

(ب) قيمة الاتجاه للفصل الرابع لعام 1993 هي :

$$T = 6040.45 - 278.01 \times 16 = 1592.29$$

وبما أن مركبة الفصل الرابع هي 129.08

إذا القيمة المتوقعة للفصل الرابع تساوي :

$$1592.29 \times \frac{129.08}{100} = 4926$$ زائرا

فلو كان عدد الزوار الفعلي 5291، فإن الخطأ في التقدير يساوي $365 = |4926-5291|$.

تمارين

1-12 : أوجد اتجاه السلسلة الزمنية التالية التي تمثل المبيعات لمدة 20 يوماً بطريقة المعدل النصفي وارسم شكل الانتشار وخط الاتجاه.

اليوم	المبيعات	اليوم	المبيعات	اليوم	المبيعات	اليوم	المبيعات
1	23	6	33	11	24	16	21
2	31	7	36	12	32	17	38
3	42	8	44	13	50	18	62
4	36	9	32	14	30	19	41
5	29	10	30	15	28	20	30

2-12 : أوجد اتجاه السلسلة الزمنية في تمرين (12-1) بطريقة المعدل المتحرك بطول 5 وأوجد مركبة التذبذب.

3-12 : إذا كانت معادلة انحدار Y على الزمن (خط اتجاه السلسلة Y) هو :

$$\hat{Y} = 732 + 0.8X$$

حيث X بالسنوات ، X = 0 تقابل عام 1989.

أ) أوجد خط الاتجاه بالأشهر.

ب) اجعل المركز X = 0 يقابل عام 1995.

جـ) اجعل خط الاتجاه الجديد بالأشهر.

4-12 : يمثل الجدول (1) في البند (12-1) عدد البواخر المحملة والمفرغة في ميناء العقبة حسب السنوات 1976 وحتى 1989. أوجد خط اتجاه عدد البواخر بطريقة المربعات الصغرى وارسم الخط على شكل الانتشار.

5-12 : الجدول التالي يمثل زوار متحف الآثار الأردني في مدينة عمان حسب الأشهر خلال عامي 1990، 1991. أوجد مركبة الفصل لكل شهر بطريقة المعدلات المتحركة بطول 12 شهراء، ثم أوجد القيم اللافصلية لعدد الزوار عام 1991.

(3) 1991 المجموع	(2) 1990 المجموع	(1) الأشهر
195	2117	كانون الثاني
430	3534	شباط
216	6318	آذار
302	5905	نيسان
330	2977	أيار
419	1114	حزيران
475	1288	تموز
1250	352	آب
790	336	أيلول
2372	1898	تشرين الأول
1177	704	تشرين الثاني
1642	1507	كانون الأول

6-12 : يمثل الجدول التالي السلسلة الزمنية التي تعطي عدد الأطفال من زوار الأماكن الأثرية في جرش حسب الأشهر خلال عامي 1988، 1989.

(3) الأطفال	(2) الأطفال	(1) الأشهر
300	200	كانون الثاني
100	80	شباط
300	300	آذار
800	200	نيسان

أيار	300	400
حزيران	300	200
تموز	606	300
آب	1300	800
أيلول	500	400
تشرين الأول	299	200
تشرين الثاني	102	200
كانون الأول	100	144
المجموع	4287	4144

(أ) أوجد المركبة الفصلية لكل شهر بطريقة النسبة إلى خط الاتجاه.

(ب) أوجد القيم اللافصلية لعام 1989.

(جـ) قدر عدد الزوار الأطفال في شهر أيار 1990 من خط الانحدار (خط الاتجاه).

(د) أوجد تقديرا لعدد الزوار الأطفال في شهر أيار 1990 آخذا بالاعتبار الاتجاه والمركبة الفصلية.

(هـ) إذا كان عدد الزوار الأطفال للأماكن الأثرية في جرش في شهر أيار 1990 هو 400 فما الخطأ في تقديرك.

7-12 : يمثل الجدول التالي : النقل الجوي للركاب على الملكية الأردنية حسب الأشهر خلال السنوات 1989-1985.

الشهر	الركاب				
	1989	1987	1989	1986	1985
كانون الثاني	95202	98495	84546	92833	90339
شباط	77941	90431	65268	67781	78458
آذار	101215	89275	72920	82096	100113

نيسان	87389	89196	82157	84582	103818
أيار	87693	83786	74841	80308	101404
حزيران	112863	106763	99573	110736	123947
تموز	146448	151198	138445	129157	141265
آب	149889	150597	148054	148930	161418
أيلول	108010	112489	111996	106637	128687
تشرين الأول	82947	94925	86179	80888	94155
تشرين الثاني	74852	80295	73892	70604	78006
كانون الأول	79556	88484	81783	77415	88684
المجموع	1204005	1225934	1119664	1131967	1290294

أ- أوجد المركبة الفصلية لكل شهر بطريقة النسبة إلى المعدل المتحرك.

ب- أوجد القيم اللافصلية لعام 1989.

8-12 : سلسلة زمنية فيها مركبة الاتجاه :

$$T = 32 + 0.2\,X$$

حيث وحدة X هي الفصل، والمركز $X = 0$ يقابل الفصل (الربع) الثالث لعام 1995.

وكانت المركبات الفصلية كما يلي :

الربع الأول	الربع الثاني	الربع الثالث	الربع الرابع
79.2	95.8	120.5	104.5

(أ) أوجد تقديرات (القيم المتوقعة) لفصول عام 2000.

(ب) إذا كانت مشاهدات عام 1996 كما يلي :

الربع الرابع	الربع الثالث	الربع الثاني	الربع الأول
36.89	42.12	34.45	26.66

أوجد القيم اللافصلية لفصول عام 1996.

إذا قدرت قيمة الربع الرابع لعام 1996 فما الخطأ في التقدير الذي حصلت عليه.

ملاحظة: المصدر للتمارين 4، 5، 6، 7 هو: النشرة الإحصائية السنوية / دائرة الإحصاءات العامة . الأردن العدد 40 لعام 1989 والعدد 42 لعام 1991.

9-12: أجب عن الأسئلة في التمرين (12-7) باستعمال الجدول (2) في البند (12-1).

10-12: أوجد اتجاه السلسلة الزمنية في مثال (1) بطريقة المعدلات المتحركة بطول 3، ثم بطريقة المربعات الصغرى. وارسم هذين الاتجاهين على شكل الانتشار للسلسلة الأصلية.

الفصل الثالث عشر
الأرقام القياسية

Index Numbers

1-13 : مقدمة

الأرقام القياسية هي طريق مختصر لوصف المتغيرات الاقتصادية. وهي كثيرا ما تستعمل لوصف التغير في الأسعار أو الكميات أو القيمة مع الزمن.

فالرقم القياسي للأسعار هو عدد واحد يرينا كيف تغيرت مجموعة كاملة من الأسعار. والرقم القياسي لتكاليف المعيشة يعطينا نسبة التغير في تكلفة الحياة بين سنة وأخرى مقاسة بتغير أسعار سلة من المواد من وقت لآخر. فمثلا لو قلنا أن الرقم القياسي لتكاليف المعيشة في الأردن بلغ 180 في عام 1995 مقارنة لعام 1990 فهذا يعني أن ما تحتاج إلى إنفاقه عائلة في عام 1995 لشراء عدد من المواد (عادة

تحدد هذه المواد عند تعريف الرقم القياسي) يزيد %80 عما كانت تحتاجه هذه العائلة عام 1990. أي أن تكاليف الحياة قد ارتفعت %80 في عام 1995 عما كانت عليه عام 1990.

ويمكن استعمال الأرقام القياسية لقياس التغير في الإنتاجية والبطالة ونمو الموارد ومعدلات الأجور وغيرها. ويستفاد من معرفة الأرقام القياسية لتكاليف الحياة في تقرير الزيادة في الرواتب والأجور وحساب راتب الضمان الاجتماعي والتقاعد.

وبمعرفة الأرقام القياسية التي تعطينا فكرة واضحة عن التغير في أسعار المواد الغذائية أو تكاليف العلاج تستطيع العائلة من تنظيم ميزانية مصروفاتها ونفقاتها لتتناسب مع دخلها.

وعادة ما نحسب :

(1) الرقم القياسي للأسعار (Price Index) الذي يقارن التغير في السعر من فترة لأخرى.

(2) الرقم القياسي للكميات (Quantity Index) الذي يقيس التغير في كمية أحد المتغيرات مع الزمن.

(3) الرقم القياسي للقيمة (Value Index) الذي يقيس التغير في القيمة المالية للمتغير. أي أنه يقيس القيمة النقدية (بالدينار مثلا) لمتغير معين. والرقم القياسي للقيمة يركب التغير في الأسعار والكميات ليعطي رقما أكثر دلالة ومعنى.

ومن أهم وجوه استخدام الأرقام القياسية تقدير الدخل الحقيقي للفرد، ويعرف الرقم القياسي للدخل الحقيقي للفرد في سنة معينة (سنة المقارنة) إلى سنة أخرى (سنة الأساس) بأنه :

$$\text{الرقم القياسي للدخل الحقيقي} = R.I. = \frac{\text{الرقم القياسي للدخل}}{\text{الرقم القياسي لتكاليف المعيشة}}$$

ويمكن الضرب بـ 100 لجعل الرقم نسبة مئوية.

فمثلا ، إذا كان الرقم القياسي لدخل الفرد عام 1995 بالنسبة لعام 1990 يساوي 160 وكان الرقم القياسي لتكاليف المعيشة لعام 1995 بالنسبة لعام 1990 يساوي 195 فإن الرقم القياسي للدخل الحقيقي : $\frac{160}{195} \times 100 = 82.05$ أي أن الدخل الحقيقي للفرد عام 95 (القوة الشرائية لدخل الفرد) أصبحت 82 بالمائة مما كانت عليه عام 1990 . سندرس في البنود التالية عدة أنواع من الأرقام القياسية.

2-13 : الأرقام القياسية البسيطة : Simple Index Number

وتعتمد هذه الأرقام في تركيبها على قيمة ظاهرة أو قيم ظواهر (متغيرات) في أزمنة أو أماكن مختلفة، ويمكن حسابها بطريقتين ونقتصر في حسابها على الأسعار :

(1) الرقم القياسي التجميعي البسيط للأسعار :

Simple Aggregate Price Index

وهي نسبة مجموع أسعار عدة سلع في سنة ما (تسمى نسبة المقارنة) إلى مجموع أسعار هذه السلع في سنة أخرى (تسمى سنة الأساس).

فلو كان لدينا الجدول التالي : حيث P_{oi} سعر السلعة **i** في سنة الأساس، P_{ni} سعر السلعة **i** سنة المقارنة.

السلعة	السعر عام 1992 (سنة الأساس)	السعر عام 1998 (سنة المقارنة)
(1)	P_{o1}	P_{n1}
(2)	P_{o2}	P_{n2}
.	.	.
.	.	.
.	.	.
(m)	P_{om}	P_{nm}

فيكون الرقم القياسي التجميعي البسيط للأسعار عام 1998 مقارنة مع أسعار 1990 هو :

$$\mathbf{I_P(a) = \frac{\sum_i P_{ni}}{\sum_i P_{oi}} \times 100}$$

ولتبسيط استعمال الرموز نتوقف عن استعمال (i) ونكتفي بالرمز لسنة الأساس بالحرف (o) وسنة المقارنة بالحرف (n) وبالتالي يكون :

تعريف (1) : الرقم القياسي التجميعي للأسعار هو :

$$\mathbf{I_P(a) = \frac{\sum P_n}{\sum P_o} \times 100}$$

ونضع (a) لتدل على التجميعي (aggregate) ونضع P لتدل على الأسعار (Price).

نصطلح على إعطاء الأرقام القياسية كنسب مئوية ولذلك ضربنا بالعدد 100.

مثال (1) : كانت الأسعار (بالفلس/كغم) لبعض المواد الاستهلاكية كما في الجدول التالي :

السلعة	السعر عام 1992	السعر عام 1998
السكر	200	300
الأرز	240	400
الشاي	1500	1800
القهوة	2200	4500
المجموع	4140	7000

احسب الرقم القياسي التجميعي البسيط للأسعار باعتبار عام 1992 سنة أساس.

الحل : الرقم القياسي التجميعي البسيط للأسعار = $\frac{\sum P_n}{\sum P_o} \times 100$

ويساوي : $\frac{7000}{4140} \times 100 = 169.1$

لاحظ أنه يجب أن تكون وحدة الأسعار هي نفسها لجميع السلع. أي أما جميعها بالفلس أو بالدينار أو بغيره.

(2) الرقم القياسي النسبي البسيط للأسعار :

Simple Relative Price Index

وهو الوسط الحسابي للأرقام القياسية للسلع. أي أننا نجد الرقم القياسي لكل سلعة ثم نجد الوسط الحسابي لهذه الأرقام القياسية.

تعريف (2) : الرقم القياسي النسبي البسيط للأسعار هو :

$$I_P(r) = \frac{1}{m}\sum \frac{P_n}{P_o}$$

حيث m هو عدد السلع.

ونضع (r) لتدل على "النسبي" relative. ونضع P لتدل على الأسعار Price.

مثال (2) : أوجد الرقم القياسي النسبي البسيط للأسعار في مثال (1) باعتبار عام 1992 سنة أساس.

الحل :

الرقم القياسي النسبي البسيط لأسعار 1998 هو :

$$I_P(r) = \frac{1}{4}\left[\frac{300}{200} + \frac{400}{240} + \frac{1800}{1500} + \frac{4500}{2200}\right] \times 100$$
$$= \frac{641.21}{4} = 160.3$$

3-13 : الأرقام القياسية المرجحة

Weighted Aggregates Index

إن من سلبيات الأرقام القياسية البسيطة أنها تعطي جميع السلع الأهمية نفسها، أي أنها لا تعطي أهمية ، أو وزنا، للتغير في سعر السلعة الأكثر استعمالا أكبر مما تعطيه للسلعة قليلة الاستعمال. فعلى سبيل المثال، الأرقام القياسية البسيطة للأسعار تعامل التغير في سعر التلفزيون أو آلة التسجيل أو الآلة الحاسبة اليدوية مثلما تعامل التغير في سعر السكر والأرز والشاي مع أن شراء السلع الثلاث الأولى يكاد يكون مرة واحدة في الفترة الزمنية الطويلة (سنة أو أكثر) بينما يكون استعمال السلع الثلاث الأخرى متكررا في اليوم الواحد أو الأسبوع الواحد.

مما سبق بات ضروريا أن نعطي لبعض السلع أهمية أكبر ووزنا أكبر مما نعطيه للسلع الأخرى. إن هذا الترجيح يعطينا الفرصة لاستعمال معلومات إضافية زيادة عن المعلومات المتعلقة بتغير الأسعار. وتبقى مسألة تحديد الوزن الذي نستعمله مع كل واحد من المتغيرات (السلع) تحت الدراسة.

وهناك العديد من الطرق التي تعين الأوزان ندرس منها :

(1) طريقة لاسبير Laspeyrs Method

وتتلخص هذه الطريقة باستعمال الكميات المستهلكة والقيمة النقدية للكميات المستهلكة في سنة الأساس كأوزان لأسعار المواد الداخلة في حساب الرقمين القياسيين التجميعي والنسبي على التوالي.

تعريف (3) : رقم لاسبير التجميعي للأسعار هو :

$$I_P(L) = \frac{\sum P_n Q_o}{\sum P_o Q_o} \times 100$$

حيث P_n = أسعار سنة المقارنة.

P_o = أسعار سنة الأساس.

Q_o = الكميات المستهلكة في سنة الأساس.

واستعملنا الرمز L ليدل على لاسبير :

تعريف (4) : رقم لاسبير النسبي للأسعار هو :

$$I_P(rL) = \sum (P_n / P_o) W_o \times 100$$

حيث : $$W_o = \frac{P_o Q_o}{\sum P_o Q_o}$$

واستعملنا الرمز rL ليدل على لاسبير النسبي .

مثال (3) : يبين الجدول التالي أسعار عدد من السلع (فلس/كغم) وكميات الاستهلاك منها بالكغم للعائلة الواحدة شهريا.

لسلعة	السعر عام 1993	كمية الاستهلاك عام 1993	السعر عام 1999	كمية الاستهلاك عام 1999	P_oQ_o	Wo
السكر	220	7	350	8	1540	$\frac{154}{2229}$
الأرز	280	10	430	12	2800	$\frac{280}{2229}$
الشاي	1700	1.5	3000	1.5	25500	$\frac{255}{2229}$
لحم الضأن	2800	5.5	4000	6.5	15400	$\frac{1540}{2229}$
	لاسبير	24		28	22290	1

أ- احسب رقم لاسبير القياسي التجميعي لأسعار 1999 باعتبار 1993 سنة أساس.

ب- احسب رقم لاسبير النسبي لأسعار 1999 باعتبار عام 1993 سنة أساس.

الحل : (أ)

$$I_P(L)=\frac{\sum P_nQ_o}{\sum P_oQ_o}\times 100=\frac{350\times 7+430\times 10+3000\times 1.5+4000\times 5.5}{220\times 7+280\times 10+1700\times 1.5+2800\times 5.5}\times 100$$
$$=\frac{33250}{22290}\times 100=149.2$$

(ب) نحسب أولا W_o ونضعه في عمود (6) في الجدول :

للسكر $$W_o=\frac{P_oQ_o}{\sum P_oQ_o}=\frac{220\times 7}{22290}=\frac{154}{2229}$$

للأرز $$W_o=\frac{P_oQ_o}{\sum P_oQ_o}=\frac{28}{2229}$$ وهكذا لبقية السلع.

نحسب الآن رقم لاسبير النسبي للأسعار وهو :

$$I_P(rL)=\left[\sum\frac{P_n}{P_o}\times W_o\right]\times 100=$$
$$\left[\frac{350}{220}\times\frac{154}{2229}+\frac{430}{280}\times\frac{280}{2229}+\frac{3000}{1700}\times\frac{255}{2229}+\frac{4000}{2800}\times\frac{1540}{2229}\right]\times 100=149.2$$

لاحظ أننا حصلنا على الجواب نفسه في (أ) وهذا صحيح بوجه عام، أي أن رقم لاسبير التجميعي للأسعار (باستعمال الكميات في سنة الأساس أوزانا) هو نفسه رقم لاسبير النسبي للأسعار (باستعمال القيمة المالية للاستهلاك في سنة الأساس أوزانا).

وبرهان ذلك :

$$I_P(rL)=\left[\sum\frac{P_n}{P_o}\times W_o\right]\times 100=$$
$$\sum\frac{P_n}{P_o}\times\frac{P_oQ_o}{\sum P_oQ_o}\times 100=\frac{\sum P_nQ_o}{\sum P_oQ_o}\times 100=I_P(L)$$

(2) طريقة باش Paasche Method

وتتلخص هذه الطريقة باستعمال الكميات المستهلكة (أو المنتجة) والقيمة النقدية للكميات المستهلكة في سنة المقارنة أوزانا لأسعار المواد الداخلة في حساب الرقمين القياسيين التجميعي والنسبي على التوالي .

تعريف (5) : رقم باش التجميعي للأسعار هو : $\mathbf{I_P(P) = \frac{\sum P_n Q_n}{\sum P_o Q_n} \times 100}$

حيث : P_n = أسعار سنة المقارنة.

P_o = أسعار سنة أساس.

Q_n = الكميات المستهلكة (أو المنتجة) في سنة المقارنة.

واستعملنا P داخل القوس لتدل على باش.

تعريف (6) رقم باش النسبي للأسعار هو :

$$\mathbf{I_P(rp) = \sum \frac{P_n}{P_o} \times W_n \times 100}$$

حيث :

$$\mathbf{W_n = \frac{P_n Q_n}{\sum P_n Q_n}}$$

مثال (4) : في مثال (3) :

أ- احسب رقم باش القياسي التجميعي لأسعار عام 1999 باعتبار عام 1993 سنة أساس.

ب- احسب رقم باش القياسي النسبي لأسعار عام 1999 باعتبار عام 1993 سنة أساس.

الحل : نعيد كتابة الجدول ونضيف عليه الأعمدة المطلوبة في الحل وسنتبع هذا النمط في حل المسائل بشكل عام.

السلعة	السعر عام 1993	كمية الاستهلاك 1993	السعر عام 1999	كمية الاستهلاك 1999	P_oQ_n	P_nQ_n	W_n
السكر	220	7	350	8	1760	2800	$\frac{280}{3846}$
الأرز	280	10	430	12	3360	5160	$\frac{516}{3846}$
الشاي	1700	1.5	3000	1.5	2550	4500	$\frac{450}{3846}$
لحم الضأن	2800	5.5	4000	6.5	18200	26000	$\frac{2600}{3846}$
					25870	38460	

(أ) رقم باش القياسي التجميعي للأسعار هو :

$$I_P(P) = \frac{38460}{25870} \times 100 = 148.67$$

(ب) رقم باش القياسي النسبي للأسعار هو :

$$100 \times \left[\frac{350}{220} \times \frac{280}{3846} + \frac{430}{280} \times \frac{516}{3846} + \frac{3000}{1700} \times \frac{450}{3846} + \frac{4000}{2800} \times \frac{2600}{3846}\right] = 149.41$$

(3) طريقة الأوزان الثابتة التجميعية Fixed Weight Aggregates Method

وتتلخص هذه الطريقة باستعمال الكميات المستهلكة (أو المنتجة) لسنة معينة باعتبارها سنة ممثلة لنمط الاستهلاك أوزانا لأسعار المواد الداخلة في حساب الرقم القياسي التجميعي. وبعبارة أخرى، لا نستعمل كميات سنة الأساس أو سنة المقارنة، بل نستعمل كميات سنة نرى أنها ممثلة لنمط الاستهلاك في المجتمع.

تعريف (7) : الرقم القياسي التجميعي المرجح للأسعار بأوزان ثابتة هو :

$$I_P(C) = \frac{\sum P_n Q}{\sum P_o Q} \times 100$$

حيث Q = كميات الاستهلاك في السنة المختارة.

نضع (C) لتدل على الوزن الثابت (Constant).

مثال (5) : في مثال (3)، لو اعتبرنا كميات الاستهلاك سنة 1995 أوزانا، وكانت هذه الكميات معطاة في الجدول كما يلي (نضيفه إلى معطيات مثال (3)).

السلعة	كمية الاستهلاك عام 1995
السكر	7.5
الأرز	13
الشاي	2
لحم الضأن	5

احسب الرقم القياسي التجميعي المرجح بأوزان ثابتة (هي كميات 1995) لأسعار 1999 باعتبار سنة 1993 سنة أساس.

الحل :

$$\mathbf{I_P(C)=\frac{\sum P_nQ}{\sum P_oQ}\times 100=}$$

$$\mathbf{=\frac{350\times 7.5+430\times 13+3000\times 2+4000\times 5}{220\times 7.5+280\times 13+1700\times 2+2800\times 5}\times 100}$$

$$\mathbf{=\frac{34215}{22690}\times 100=150.79}$$

(4) الرقم الأمثل لفشر Fisher's Ideal Index.

تعريف (8) :

الرقم الأمثل لفشر = $\sqrt{\text{رقم لاسبير} \times \text{رقم باش}}$

أي $$\mathbf{I_P(F)=\sqrt{I_P(L)\ \ I_P(P)}}$$

من المثالين (3)، (4) نجد أن الرقم الأمثل لفشر يساوي :

$$\mathbf{I_P(F)=\sqrt{149.2\times 148.67}=148.93}$$

4-13 : الأرقام القياسية ذات الأساسي المتحرك :

Index Numbers With Moving Base

إن تحديد سنة الأساس عامل مهم في تركيب الأرقام القياسية، وقد استعملنا في البنود السابقة سنة محددة كأساس ثابت، ومن عيوب هذه الطريقة أنه إذا كانت المدة بين سنة الأساس وسنة المقارنة طويلة نسبيا فإن الرقم القياسي لا يعبر تعبيرا صحيحا عن التطورات التي تنشأ خلال هذه المدة. فمثلا إذا حسبنا الرقم القياسي للأسعار أو نفقات المعيشة فقد تدخل سلع جديدة وقد تختفي سلع أخرى وقد يحدث تغير في نمط الحياة وفي السلوك الاستهلاكي للمجتمع وقد يحدث تغير كبير جدا في أسعار السلع في فترة قصيرة جدا بسبب بعض التشريعات التي تسنها الدولة. ولذلك فإن هذا الرقم القياسي لا يعبر تعبيراً صحيحاً عن التطورات الناشئة. فعلى سبيل المثال انتشر في الأعوام القليلة الماضية مشاهدة القنوات الفضائية التي تحتاج إلى أجهزة معينة وانتشرت خدمة الانترنت والهواتف النقالة وغيرها مما أثر كثيرا في عدد السلع التي يتدخل في تركيب الرقم القياسي، أضف إلى ذلك على سبيل المثال تخفيض جمارك السيارات الذي يؤدي إلى تغيير الرقم القياسي الذي يُدخل السيارة في تركيبه.

ولعلاج هذه المشاكل نستعمل طريقة غير مباشرة تؤدي للمقارنة وذلك بتكوين أرقام قياسية للفترات المتلاحقة بحيث تكون كل فترة أساسا للفترة التي تليها مباشرة، وبضرب تلك الأرقام في بعضها البعض

نحصل على الرقم القياسي المطلوب، ويسمى هذا الأسلوب بأسلوب الأساس المتحرك ويستعمل لمقارنة الحاضر بالماضي القريب وليس بالماضي البعيد.

مثال (7) : يعطي الجدول التالي أسعار ثلاث سلع استهلاكية بين سنتي 1992، 1998.

السلعة	سعر 92	سعر 93	سعر 94	سعر 95	سعر96	سعر 97	سعر 98
السكر	200	200	250	250	300	320	350
الأرز	180	180	220	250	380	400	420
الخبز	75	85	85	200	200	150	150

أوجد الرقم القياسي النسبي بأساس متحرك.

الحل : نجد الأرقام القياسية دون أن نجعلها نسبا مئوية، أي لا نضرب بـ 100. الرقم القياسي النسبي لأسعار 93 بالنسبة لعام 92 هو :

$$\frac{1}{3}\left[\frac{200}{200}+\frac{180}{180}+\frac{85}{75}\right]=1.04$$

الرقم القياسي النسبي لأسعار 94 بالنسبة لعام 93 هو :

$$\frac{1}{3}\left[\frac{250}{200}+\frac{220}{180}+\frac{85}{85}\right]=1.16$$

الرقم القياسي النسبي لأسعار 95 بالنسبة لعام 94 هو :

$$\frac{1}{3}\left[\frac{250}{250}+\frac{250}{220}+\frac{200}{85}\right]=1.5$$

الرقم القياسي النسبي لأسعار 96 بالنسبة لعام 95 هو :

$$\frac{1}{3}\left[\frac{300}{250}+\frac{380}{250}+\frac{200}{200}\right]=104$$

الرقم لقياسي النسبي لأسعار 97 بالنسبة لعام 96 هو :

$$\frac{1}{3}\left[\frac{320}{300}+\frac{400}{380}+\frac{150}{200}\right]=0.96$$

الرقم لقياسي النسبي لأسعار 98 بالنسبة لعام 97 هو :

$$\frac{1}{3}\left[\frac{350}{320}+\frac{420}{400}+\frac{150}{150}\right]=1.05$$

فيكون الرقم القياسي النسبي لأسعار 98 بالنسبة لعام 92 على الأساس المتحرك يساوي :

$$1.04 \times 1.16 \times 1.5 \times 1.24 \times 0.96 \times 1.05 \times 100 = 226.18$$

بينما الرقم القياسي النسبي لأسعار 98 بالنسبة لعام 92 على الأساس الثابت يساوي :

$$\frac{1}{3}\left[\frac{350}{200}+\frac{420}{180}+\frac{150}{75}\right]\times 100=202.78$$

5-13 تغيير سنة الأساس والتوصيل لسلسلة أرقام قياسية

Shifting the base of an index series and splicing

(1) تغيير سنة الأساس

نحتاج في بعض الأحيان لتغيير سنة الأساس لسلسلة من الأرقام القياسية ولا نستطيع إعادة حساب الأرقام القياسية بالنسبة للأساس الجديد من البيانات الأصلية لأنها عادة لا تكون متوفرة لدينا. لذلك نعمد إلى طريقة سهلة كما يلي :

مثال (8) : افرض أن لديك سلسلة أرقام قياسية على اعتبار 1990 سنة أساس :

السنة	87	88	89	90	91	92	93	94	95	96	97
الرقم القياسي	80	86	95	100	110	140	142	155	170	181	192

وأردنا جعل 1987 سنة أساس، فالحل يكون بقسمة كل رقم قياسي على $\frac{80}{100}$ أي نقسم على :

الرقم القياسي لسنة الأساس الجديدة

الرقم القياسي لسنة الأساس الأصلية

فنجد :

السنة	87	88	89	90	91	92	93	94	95	96	97
الرقم القياسي	100	107.5	118.75	125	137.5	175	154	193.75	212.5	226.22	240

وبالمثل ، لو أردنا جعل عام 1994 سنة أساس لحصلنا على سلسلة الأرقام القياسية بقسمة كل رقم قياسي في السلسلة على $1.55 = \frac{155}{100}$.

(2) التوصيل Splicing

من المعلوم أننا نحسب الأرقام القياسية بشكل مستمر ولكن ربما باستعمال سنة أساس محددة لكل فترة زمنية. إذا أردنا تجميع أو توصيل هذه السلاسل من الأرقام القياسية لتكون سلسلة واحدة باعتبار أساس واحد فإننا نستعمل أسلوب تغيير سنة الأساس.

مثال (9) : لديك سلسلتان من الأرقام القياسية سنة الأساس في الأولى عام 1988 وتنتهي عام 1993 وسنة الأساس في الثانية عام 1993. اجعل منهما سلسلة واحدة على اعتبار 1988 سنة أساس.

السنة	88	89	90	91	92	93	94	95	96	97	98
الرقم القياسي	100	110.5	116.3	127.9	162.8	165.1 100	(180.3) 109.2	(197.6) 119.7	(210.5) 127.5	(223.2) 135.2	(251.0) 152.1

الحل : نضرب الرقم القياسي للأعوام التي أساسها سنة 1993 بالكسر $\frac{165.1}{100}$ ونكتب الإجابة في الخانات المخصصة لها ونضعها بين أقواس لنبين أن هذه الأرقام حسبت لتوصيل السلسلة ولم تكن موجودة باعتبار سنة الأساس الأصلية. فمثلا الرقم القياسي لعام 94 يساوي

$180.3 = 109.2 \times \frac{165.1}{100}$ وهكذا نحسب الأرقام القياسية الأخرى فنحصل على سلسلة أرقام قياسية على أساس 1988.

6-13 الأرقام القياسية للكميات : Quantity Index Numbers

اقتصر اهتمامنا في البنود السابقة على حساب الأرقام القياسية للأسعار وبما أن كميات الإنتاج أو الاستهلاك تتغير مع الزمن أصبح ضروريا أن نقارن التغير في الكميات من فترة زمنية لأخرى

وبالتالي نجد الرقم القياسي للكمية في فترة زمنية بالنسبة لفترة زمنية أخرى نعتبرها الأساس. وحساب الأرقام القياسية للكميات يكون بنفس الطريقة التي حسبنا فيها الأرقام القياسية للأسعار مع تبديل الأدوار. فنحصل على التعاريف التالية :

1- الرقم القياسي البسيط للكميات هو :

$$I_q(a) = \frac{\sum Q_n}{\sum Q_o} \times 100$$

2- الرقم القياسي النسبي البسيط للكميات :

$$I_q(r) = \frac{1}{m} \sum \frac{Q_n}{Q_o} \times 100$$

3- رقم لاسبير التجميعي للكميات :

$$I_q(L) = \frac{\sum Q_n P_o}{\sum Q_o P_o} \times 100$$

4- رقم لاسبير النسبي للكميات هو :

$$I_q(rL) = \sum \left(\frac{Q_n}{Q_o} \right) \times T_o \times 100$$

حيث

$$T_o = \frac{Q_o P_o}{\sum Q_o P_o}$$

مع ملاحظة أن (4) يساوي (3).

5- رقم باش التجميعي للكميات :

$$I_q(P) = \frac{\sum Q_n P_n}{\sum Q_o P_n} \times 100$$

6- رقم باش النسبي للكميات :

$$I_q(rP) = \sum \frac{Q_n}{Q_o} \times T_n \times 100$$

حيث $$T_n = \frac{Q_n P_n}{\sum Q_n P_n}$$

7- الرقم القياسي التجميعي المرجح للكميات بأوزان ثابتة :

$$I_q(C) = \frac{\sum Q_n P}{\sum Q_o P} \times 100$$

حيث P = أسعار السلع في السنة المختارة.

8- رقم فيشر الأمثل للكميات :

$$I_q(F) = \sqrt{I_q(L)\ \ I_q(P)}$$

أي : رقم لاسبير × رقم باش

وتحسب الأرقام القياسية للكميات بأساس متحرك ويتغير أساس سلسلة الأرقام القياسية ويتم توصيلها بنفس الطريقة للأرقام القياسية للأسعار.

مثال (10) : في مثال (3) :

(أ) احسب $I_q(a)$, $I_q(r)$, $I_q(L)$, $I_q(P)$ حيث سنة الأساس 1993.

الحل : بالرجوع إلى التعاريف نجد :

$$I_q(a) = \frac{8+12+1.5+6.5}{7+10+1.5+5.5} \times 100 = \frac{28}{24} \times 100 = 116.7$$

$$I_q(r) = \frac{1}{4}\left[\frac{8}{7} + \frac{12}{10} + \frac{1.5}{1.5} + \frac{6.5}{5.5}\right] \times 100 = 113.1$$

$$I_q(L) = \frac{8\times 220 + 12\times 280 + 1.5\times 1800 + 6.5\times 2800}{7\times 220 + 10\times 280 + 1.5\times 1700 + 5.5\times 2800} \times 100$$
$$= \frac{25870}{22290} \times 100 = 116.1$$

$$I_q(P) = \frac{8\times 350 + 12\times 430 + 1.5\times 3000 + 6.5\times 4000}{7\times 350 + 10\times 430 + 1.5\times 3000 + 5.5\times 4000}\times 100$$

$$= \frac{38460}{33250}\times 100 = 115.7$$

(ب) احسب رقم فيشر الأمثل للكميات :

$$I_q(F) = \sqrt{116.1\times 115.7} = 115.9$$

7-13 الأرقام القياسية للقيمة : Value Index Numbers

إضافة إلى الأرقام القياسية للأسعار والكميات نحتاج بعض الأحيان لاستعمال الأرقام القياسية للقيمة النقدية للسلع الاستهلاكية المستعملة والداخلة في حساب الرقم القياسي ؛ والقيمة النقدية للسلع في فترة المقارنة (n) هي :

$$V_n = \sum P_n Q_n$$

والقيمة النقدية للسلع في فترة الأساس (o) هي :

$$V_o = \sum P_o Q_o$$

ويعرف الرقم القياسي البسيط للقيمة بأنه :

$$I_v = \frac{V_n}{V_o} = \frac{\sum Q_n P_n}{\sum Q_o P_o}$$

مثال (11) : في مثال (4) احسب الرقم القياسي للقيمة.

الحل :

$$I_v = \frac{38460}{22290}\times 100 = 172.5$$

حيث البسط والمقام قد حسبت قيمتاهما من قبل.

8-13 اختبار الأرقام القياسية : Tests of Index Numbers

تختلف الأرقام القياسية من حيث كونها نسبية أو تجميعية ومن حيث الترجيح أو عدمه ومن حيث استعمال الأساس الثابت أو الأساس المتحرك. ومن الطبيعي أن ينشأ التساؤل عن أي نوع من الأرقام القياسية نستعمل. والجواب على ذلك يعتمد على الغرض من بناء الرقم القياسي ومن توفر البيانات وخاصة الأسعار والكميات في سنة الأساس والسنوات المقارنة، فإذا توفرت الأسعار فقط فلا نستطيع

الترجيح بأوزان الكميات وكذلك إذا توفرت الكميات فقط فلا نستطيع الترجيح بأوزان الأسعار وهكذا.
أما تركيب الرقم القياسي فيعتمد على بعض الاختبارات الإحصائية:

(1) اختبار الانعكاس في الأساس : Time reversal Property

وينص هذا الاختبار على أن حاصل ضرب الرقمين القياسيين المتبادلين يجب أن يساوي الواحد الصحيح. أي إذا حسبنا رقما قياسيا باعتبار الأساس فترة معينة أو سنة محددة ثم حسبنا رقما قياسيا بمبادلة دور سنة المقارنة وسنة الأساس وضربنا هذين الرقمين في بعضها البعض بعد قسمة كل منهما على 100 يكون الجواب 1.

من الواضح أن الرقم القياسي التجميعي البسيط للأسعار بعد القسمة

على 100 (وكذلك للكميات) يحقق اختبار الانعكاس في الأساس، حيث :

$$\mathbf{I_P^{(2)}(a) = \frac{\sum P_o}{\sum P_n}} \quad , \quad \mathbf{I_P^{(1)}(a) = \frac{\sum P_n}{\sum P_o}}$$

إذا :

$$\mathbf{I_P^{(1)}(a) \times I_P^{(2)}(a) = \frac{\sum P_n}{\sum P_o} \times \frac{\sum P_o}{\sum P_n} = 1}$$

مثال (12) : في مثال(1) احسب الرقم القياسي التجميعي البسيط لأسعار 1992 باعتبار عام 1998.
سنة أساس؛ واختبر فيما إذا كان يحقق اختبار الانعكاس في الأساس.

الحل :

$$\mathbf{I_P^{(2)}(a) = \frac{4140}{7000}}$$
$$\mathbf{= 0.591}$$

ومن مثال (1) كان :

$$\mathbf{I_P^{(1)}(a) = 1.691}$$

إذا :

$$\mathbf{I_P^{(1)}(a) \times I_P^{(2)}(a) = 1.691 \times 8.591 \cong 1}$$

مثال (13) : في مثال (2) هل الرقم القياسي النسبي البسيط يحقق اختبار الانعكاس في الأساس .

الحل : نحسب الأرقام القياسية دون أن نجعلها نسبا مئوية :

(1) باعتبار سنة 1992 أساساً فإن الرقم القياسي $= \mathbf{I_P^{(1)}(r) = 1.603}$.

(2) باعتبار سنة 1998 أساساً فإن الرقم القياسي :

$$\mathbf{I_P^{(2)}(r) = \frac{1}{4}\left[\frac{200}{300} + \frac{240}{400} + \frac{1500}{1800} + \frac{2200}{4500}\right] = 0.647}$$
$$\mathbf{= 1.603 \times 0.647 = 1.04}$$

الجواب قريب من 1 وبالتالي في هذا المثال فالرقم القياسي النسبي تقريبا يحقق اختبار الانعكاس في الأساس وهذا ليس بالضرورة دائما لهذا الرقم القياسي.

عندما يكون حاصل ضرب الرقمين القياسيين المتبادلين أكبر من 1 يكون الرقم القياسي منحازا للأعلى وعندما يكون حاصل الضرب أصغر من 1 يكون الرقم القياسي منحازا للأدنى. وأفضل الأرقام القياسية من وجهة النظر هذه هو الذي يحقق اختبار الانعكاس في الأساس.

(2) اختبار الانعكاس في العامل : Factor reversal Property

وينص هذا الاختبار على أن : الرقم القياسي للأسعار × الرقم القياسي للكميات = الرقم القياسي للقيمة.

وعند تطبيق هذا الاختبار على الأرقام القياسية المختلفة وجد أنها لا تحققه بشكل عام ولكن بعضها يحققه في حالات خاصة. أما الرقم القياسي الأمثل فإنه يحقق هذا الاختبار ، أي أن :

$$\mathbf{I_P(F) \times I_q(F) = I_v(F)}$$

مثال (14) : في مثال (3) ومثال (11) ، هل يحقق الرقم القياسي التجميعي اختبار الانعكاس في العامل ؟

الحل : الرقم القياسي التجميعي للأسعار بالنسبة لعام 1993 يساوي :

$$\mathbf{I_P(a) = \frac{7000}{4140} \times 100 = 169.1}$$

الرقم القياسي التجميعي البسيط للكميات بالنسبة لعام 1993 يساوي :

$$\mathbf{I_q(a) = \frac{28}{24} \times 100 = 116.7}$$

الرقم القياسي التجميعي البسيط للقيمة بالنسبة لعام 1993 يساوي :

$$I_v = \frac{38460}{22290} \times 100 = 172.5$$

هل : $I_P(a) \times I_q(a) = I_v$ بعد قسمة كل واحد منها على 100.

$$\frac{169.1}{100} \times \frac{116.7}{100} = 1.97 \neq 1.725$$

إذاً الرقم القياسي التجميعي لا يحقق اختبار الانعكاس في العامل، وبما أن 1.97 أكبر من 1.72 فالرقم ا لقياسي منحاز للأعلى.

مثال (15) : في مثال (3) وما يعتمد عليه مثال (6)، مثال (10) ومثال (11) هل يحقق رقم فيشر الأمثل اختبار الانعكاس في العامل.

الحل :

من مثال (6) : $I_P(F) = \sqrt{149.2 \times 148.67} = 148.93$

ومن مثال (10) : $I_q(F) = \sqrt{116.1 \times 115.7} = 115.9$

من مثال (11) : $I_v = 172.54$

هل : $$\frac{I_P(F)}{100} \times \frac{I_q(F)}{100} = \frac{I_v}{100}$$

$$1.489 \times 1.159 = 1.726 \approx 1.7254$$

إذا يحقق رقم فيشر الأمثل اختبار الانعكاس في العامل.

يجب ملاحظة أنه لدى إجراء اختبارات الانعكاس في الأساس والانعكاس في العامل يجب تطبيق كل اختبار على الأرقام القياسية كنسب وليس كنسب مئوية، أي نقسم كل رقم قياسي على 100 قبل إجراء الاختبار وينطبق هذا على حساب الرقم القياسي بأساس متحرك.

تمارين

1-13 اختبر نفسك، ما معنى ما يلي :

$$I_q(a), I_P(r), I_P(a), I_q(L), I_P(rL), I_P(L), I_q(r),$$
$$I_q(rP), I_q(P), I_q(rp), I_P(P), I_v, I_q(F), I_P(F)$$

2-13 : يعطي الجدول التالي معدل وفيات الأطفال الرضع (لكل 1000 طفل ولد حيا) في الأردن[1] .

السنة	1950	60	70	80	88	1990
المعدل	162	151	87	70	49	45

أ- أوجد الرقم القياسي لمعدل الوفيات لجميع السنوات في الجدول باعتبار سنة 1950 أساسا.

ب- أوجد الرقم القياسي لمعدل الوفيات عام 1990 باعتبار سنة 1970 سنة أساس.

جـ- أوجد الرقم القياسي لمعدل الوفيات عام 1990 بالنسبة لسنة 1950 بأسلوب الأساس المتحرك.

3-13 : من مصدر التمرين 13-2 لديك البيانات التالية :

السنة	1950	60	70	80	88	1990
توقعات الحياة عند الولادة بالسنوات	43.2	48.2	56.6	64.2	68	69

أجب عن الأسئلة كما في التمرين (13-2).

---

[1] المصدر : النشرة السكانية / اللجنة العامة للجنة الوطنية للسكان، العدد الأول السنة الأولى 1990.

4-13 : لديك البيانات التالية :

كمية الاستهلاك 98	كمية الاستهلاك 95	كمية الاستهلاك 1993	سعر عام 1998	سعر عام 1995	سعر عام 1993	السلعة
10	12	10	400	240	200	الأرز
3	3	2	2800	2500	2100	الزيت
8	7	6	4000	3500	3200	لحم الضأن
4	2	2	1700	1200	1000	السمك
9	11	10	1300	950	650	الدجاج

وحدات الأسعار : فلس/كغم. وكمية الاستهلاك : كغم بالشهر.

احسب : (أ) $I_P(a)$ لعام 98 بالنسبة لعام 1993.

(ب) $I_q(a)$ لعام 98 بالنسبة لعام 1993.

(جـ) I_v لعام 98 بالنسبة لعام 95 وكذلك بالنسبة لعام 93 مع ذكر معنى I_v.

(د) $I_q(a)$ ، $I_P(a)$ لعام 98 بالنسبة لعام 1995.

5-13 في تمرين (13-4).

أ- احسب الرقم القياسي النسبي لأسعار 1998 بالنسبة لأسعار 93 بأسلوب الأساس المتحرك.

ب- احسب الرقم القياسي النسبي لكميات 98 بالنسبة لكميات 93 بأسلوب الأساس المتحرك.

جـ- احسب الرقم القياسي التجميعي لقيمة الانفاق عام 1998 بالنسبة لعام 93 بأسلوب الأساس المتحرك.

6-13 : في تمرين (13-4) احسب :

أ- $L_p(L)$ ، $L_p(P)$ حيث سنة الأساس 1993، وسنة المقارنة 1998، ما معنى كل من الرقمين السابقين ؟

ب- $L_q(L)$ ، $L_q(P)$ حيث سنة الأساس 1993، وسنة المقارنة 1998.

7-13 : من نتائج التمارين (13-4) - (13-6) أي من الأرقام القياسية التالية :

أ- منحاز للأعلى ، منحاز للأسفل ، غير منحاز.

ب- يحقق اختبار الانعكاس في الأساس.

(1) الرقم القياسي النسبي البسيط لأسعار 1998 بالنسبة لعام 1995.

(2) الرقم القياسي النسبي البسيط لكميات 98 بالنسبة لعام 95.

(3) الرقم القياسي التجميعي البسيط لأسعار 98 بالنسبة عام 93.

(4) الرقم القياسي التجميعي البسيط لكميات 98 بالنسبة لعام 93.

(5) الرقم القياسي ا لتجميعي البسيط للقيم لعام 98 بالنسبة لعام 95 ثم لعام 98 بالنسبة لعام 93.

8-13 : في تمرين (13-4) وتمرين (13-6) :

(أ) احسب الرقم القياسي للقيم I_v لعام 1998 بالنسبة لعام 1993.

(ب) واحسب رقم فيشر الأمثل للأسعار $L_p(F)$ ورقم فيشر الأمثل للكميات $L_q(F)$.

(جـ) هل يحقق رقم فيشر الأمثل اختبار الانعكاس في العامل ؟

9-13 : لديك سلسلة من الأرقام القياسية تنتهي عام 1990 وأخرى تبدأ من عام 1990 كما في الجدول :

العام	85	86	87	88	89	90	91	92	93	94	95	96	97
الرقم القياسي	100	110	127	121	135	147 100	120	140	152	170	182	190	195

أ - أوصل السلسلتين (اجعلهما سلسلة واحدة) باعتبار سنة الأساس 1985.

ب- بعد وصل السلسلتين اجعل سنة الأساس 1992 واكتب السلسلة الناتجة.

13-10 : إذا كان الرقم القياسي لعدد سكان الأردن عام 1995 بالنسبة للسكان عام 1989 هو 131.72 .

أ- ما معنى هذا الرقم ؟

ب- إذا كان عدد سكان محافظة العاصمة عام 1989 يساوي 1297100 نسمة فما تقديرك لسكان العاصمة عام 1995 إذا افترضنا نمو السكان في العاصمة مثل نموه في الأردن ؟

حـ- إذا كان عدد سكان محافظة اربد عام 1995 يساوي 992400 فما هو تعداد سكان محافظة اربد عام 1989 ؟

13-11 : كان راتب موظف عام 1992 يساوي 350 دينارا وفي عام 1999 أصبح 570 دينارا فإذا كان الرقم القياسي لتكاليف المعيشة لعام 1999 بالنسبة لعام 1992 هو 210 فما القوة الشرائية لدخل الموظف في عام 99 مقارنة بعام 1992.

الفصل الرابع عشر
الإحصاءات الحيوية

Vital Statistics

1-14 : البيانات الديموغرافية Demographic Data

كما ذكرنا في الفصل الأول فإن كلمة إحصاء Statistics كانت تستعمل في الماضي لتدل على البيانات التي تجمعها الدولة State عن مجتمعها. أما في الوقت الحاضر فإننا نستعمل البيانات الديمغرافية لتدل على البيانات التي نجمعها عن المجتمع وعلى وجه الخصوص، حجم المجتمع وتركيبه، بيانات الإحصاءات الحيوية التي تحدث في المجتمع مثل (الولادات، الوفيات، عقود الزواج، حالات الطلاق) وغيرها.

وبالإضافة إلى تعداد أفراد المجتمع أصبحت الدول بحاجة ماسة إلى الإحصاءات الاجتماعية والمؤشرات الاجتماعية وأصبح النهوض بهذه الإحصاءات والعناية بها وحفظ السجلات عنها أمرا ضروريا لكل دولة نامية أو متقدمة.

فبالإضافة إلى تعداد السكان فإن السجلات الحديثة للمسح السكاني تشمل بيانات متعددة عن المجتمع مثل العمر، الجنس ، العنصر، الحالة الاجتماعية، الوظيفة وغيرها. وكذلك فالتعدادات تجرى على الأفراد في منطقة محددة في زمن محدد.

وتستعمل المعلومات المتوفرة من مسوحات الوقت الحاضر في تطبيقات كثيرة لدى الحكومات وفي الصناعة والتخطيط الصحي والتعليم. وإضافة إلى المعلومات عن حجم المجتمع وسماته، فإن البيانات المسجلة عن الحوادث الحيوية ذات أهمية بالغة. فعدد الولادات، والوفيات، وعقود الزواج وحالات الطلاق يمكن الحصول عليها من السجلات الرسمية في الدولة حيث تسجل هذه الحالات في فترة زمنية محددة مثل السنة. إن الإحصاءات عن الوفاة والمرض في مجتمع ما ذات أهمية كبرى في التخطيط لمراقبة الأمراض والحد منها.

إن مثل هذه البيانات يمكن الحصول عليها من السجلات الرسمية وهناك بيانات عن الحالات المرضية وانتشارها ذات نطاق واسع تقوم بجمعها هيئات ومؤسسات متعددة. إن جميع موضوعات البيانات المذكورة آنفا تعتبر جزءا من الإحصاءات الاجتماعية.

هناك تعاريف عديدة للإحصاءات الاجتماعية نستخلص منها بأن "الإحصاءات الاجتماعية في مجموعها ليست هدفا في حد ذاتها، بل هي عبارة عن معلومات في صورة رقمية تستخدم في تركيب مؤشرات

يطلق عليها اسم مؤشرات اجتماعية تصور كل ما يتعلق بالإنسان باعتباره هدفا ووسيلة سواء من الناحية المادية أو المعنوية" [1].

وهناك تسميات عديدة يقصد بها المؤشرات الاجتماعية منها :

مؤشرات الرفاه الوطني National Welfare Indicators

مؤشرات الرفاه الاجتماعي Social Welfare Indicators

قياسات اجتماعية Social Measurements

وغيرها كثير، إلا أن تعبير "المؤشرات الاجتماعية Social Indicators" أصبح أكثر شيوعا من غيره وذلك بعد أن نشر R. A. Bouer كتابه تحت عنوان : Social Indicators.

وهناك جهود اجتهادات عديدة لتعريف المؤشرات الاجتماعية [2] منها :

1- اجتهادات ريموند .أ. بوير R. A. Bouer.

2- اجتهادات دفيد س. روبرت D.C. Robert .

3- اجتهادات منظمة التعاون الاقتصادي والتنمية OECD.

4- اجتهادات معهد الأمم المتحدة لبحوث التنمية الاجتماعية UNRISD.

وعلى سبيل المثال ، فاجتهاد بوير يتلخص في : "المؤشرات الاجتماعية هي إحصاءات وسلاسل إحصائية وجميع أنواع الدلائل التي تساعدنا لتقويم أين نحن الآن وإلى أين نسير في ما يتعلق بقيمنا وأهدافنا تقويم البرامج وتأثيراتها"[3]. أي أن المؤشرات الاجتماعية عبارة "عن نظام للمؤشرات الإحصائية يعاون ببساطة وبوضوح في صورة رقمية على قياس كافة المجالات ذات العلاقة بحياة الإنسان في نطاق برامج التنمية"[4]. أما روبرت فيلخص التعريف كما يلي : "المؤشرات الاجتماعية هي قياسات أو إحصاءات اجتماعية مختارة[5]. وليس في مجال هذا الفصل الخوض في الاجتهادات الكثيرة لتعريف المؤشرات لاجتماعية، إلا أنه يجب الأخذ بعين الاعتبار أن هذه التعاريف.

تتضمن :

---

[1] النهوض بالإحصاءات الاجتماعية، مدخل فعال لترشيد العمل الإحصائي العربي. منشورات مجلس الوحدة الاقتصادية العربية / الأمانة العامة، عمان/ المملكة الأردنية الهاشمية، تموز 1983.

[2] النهوض بالإحصاءات الاجتماعية ـ مصدر سابق.

[3][4] ، [5] النهوض بالإحصاءات الاجتماعية. مصدر سابق.

1- مجالات استخدام المؤشرات الاجتماعية على مستوى مرحلي.

2- الصفات الواجب توفرها في المؤشرات الاجتماعية.

3- بعض الصيغ التي يتم بموجبها تركيب المؤشرات الاجتماعية.

ويهمنا في هذا المجال أن الصيغ الشكلية التي تأخذها لمؤشرات الاجتماعية قد صنفت إلى مجموعات ثلاث :

1- مؤشرات في صيغة نسبة مئوية :Percentage type

2- مؤشرات في صيغة انفرادية : Per Capita type

3- مؤشرات هيكلية Structural Indicators

وقد قامت هيئة من الأمم المتحدة ومنظماتها المختلفة مثل معهد الأمم المتحدة لبحوث التنمية الاجتماعية UNRISD، منظمة الأمم المتحدة للتربية والثقافة والعلوم UNESCO، منظمة الصحة العالمية WHO وصندوق الأمم المتحدة لرعاية الطفولة UNICEF، وغيرها مثل السوق الأوروبية المشتركة EEC واللجنة الاقتصادية لغربي آسيا ECWA بجهود كبيرة ومثمرة في تأطير المؤشرات الاجتماعية وجمع البيانات لتركيبها وتحديد مجالات استخدامها.

وبالرغم من الاختلاف في التفاصيل وعدد المؤشرات الاجتماعية لكل موضوع فإن معظم الهيئات واللجان مثل اللجنة الاقتصادية لغرب آسيا ECWA ومنظمة التعاون الاقتصادي والتنمية OECD، والسوق الأوروبية المشتركة EEC، قد اعتمدت الموضوعات الرئيسة التالية للمؤشرات الاجتماعية وهي :

1- السكان.

2- العائلة والأسرة والحالة المعيشية.

3- النشاطات التعليمية والخدمات التربوية.

4- الأعمال المأجورة وخدمات التوظيف والعاطلون عن العمل.

5- توزيع الدخل والاستهلاك والتراكم.

6- الضمان الاجتماعي وخدمات الانعاش الاجتماعي.

7- الصحة العامة والخدمات الصحية والتغذية.

8- الإسكان وبيئته.

9- النظام العام والأمن.

10- توزيعات الأوقات.

11- وقت الفراغ والثقافة.

ولكل موضوع رئيس من هذه الموضوعات أقسام فرعية وعدد من المؤشرات وسنقصر موضوعنا على موضوع "الإحصاءات الحيوية" وهو قسم من موضوع الصحة العامة والخدمات الصحية والتغذية المتعلق بالحياة والوفاة والولادة والخصوبة والمرض.

هناك تعبيران هامان لوصف الإحصاءات الحيوية وهما النسبة Ratio والمعدل Rate.

عند المقارنة بين المجتمعات (حسب تقسيمات جغرافية معينة أو أزمنة معينة) لا يكفي معرفة عدد تكرار حادث معين كالوفاة أو الولادة ولكن الأكثر أهمية هو معرفة معدل حدوث ذلك الحادث أو نسبة حدوثه.

إذا كان a = تكرار وقوع حادث ما خلال فترة زمنية محددة.

(a + b) = عدد الأفراد الذين كان من الممكن تعرضهم لهذا الحادث خلال الفترة الزمنية ذاتها، وليكن K أحد الأعداد 1، 10 ، 100 ، 1000 أو 10000 وهكذا، فإن :

$\frac{a}{(a+b)} \times K$ يسمى معدل وقوع هذا الحادث لكل K من الأفراد.

ويسمى العدد K الأساس وسنصطلح استعمال K = 1000 في البنود القادمة إلا إذا أعطيناها قيمة أخرى.

أما النسبة فهي مقدار على النحو $\left(\frac{c}{d}\right)K$ حيث أنه ليس ضروريا أن تكون c جزءا من d.

مثال (1) :

إذا كان عدد سكان مدينة 45000 نسمة في منتصف العام وبلغ عدد الوفيات في تلك المدينة في تلك السنة 700، فما هو معدل الوفاة.

من الواضح أن : عدد الوفيات = 700 = a.

وعدد السكان = 45000 = (a + b) إذا معدل الوفاة يساوي :

$$\frac{700}{45000} \times 1000 = 15.5$$

أي أن الوفيات تحصل بمعدل 15.5 لكل 1000 نسمة.

مثال (2) :

عدد سكان مدينة 60000 وفيها مستشفى حكومي واحد سعته 100 سرير وعدد الأطباء في المستشفى 24 وعدد الممرضات 70. فما هي النسب التالية : عدد الأطباء للسكان ، عدد الأسرة للسكان، عدد الممرضات للسكان.

الحل : $$\frac{24}{60000} \times 1000 = 0.4$$

إن نسبة الأطباء إلى السكان هي 0.4 لكل ألف.

$$\frac{100}{60000} \times 1000 = 1.67$$

إن نسبة الأسرة إلى السكان هي 1.67 لكل ألف.

$$\frac{70}{60000} \times 1000 = 11.17$$

إن نسبة الممرضات إلى السكان هي 1.17 لكل ألف.

مثال (3) : جاء في النشرة السكانية / العدد الأول السنة الأولى 1990 / الأمانة العامة للجنة الوطنية للسكان / عمان / الأردن أن :

عدد السكان عام 1989	3111000
ذكور	1627100
إناث	1483900

نسبة الأطباء إلى إجمالي السكان 1 : 545

نسبة أطباء الأسنان إلى إجمالي السكان 1 : 3250

نسبة الطلبة على مقاعد الدراسة : 31%

احسب : نسبة الذكور إلى إجمالي السكان.

نسبة الإناث إلى إجمالي السكان.

نسبة الذكور إلى الإناث.

عدد كل من : الأطباء ، أطباء الأسنان ، الطلبة على مقاعد الدراسة.

الحل : نسبة الذكور إلى إجمالي السكان تساوي :

$$\frac{1627100}{3111000} = 0.253 = 52.3\%$$

نسبة الإناث إلى إجمالي السكان تساوي :

$$\frac{1483900}{3111000} = 0.477 = 47.7\%$$

نسبة الذكور إلى الإناث تساوي :

$$\frac{1627100}{1483900} = 1.096$$

إي 1.096 : 1

ونسبة الإناث إلى الذكور = 1 : 1.096

عدد الأطباء = نسبة الأطباء إلى إجمالي السكان × عدد السكان ويساوي :

$$\frac{1 \times 3111000}{545} = 3708$$

عدد أطباء الأسنان يساوي :

$$\frac{1 \times 3111000}{3250} = 957$$

عدد الطلبة على مقاعد الدراسة =

$$\frac{31 \times 3111000}{100} = 964410$$

2-14 إحصاءات الوفيات : Mortality Statistics

تعبر معدلات الوفاة عن التكرارات النسبية لحدوث الوفاة ضمن مجتمع معين خلال فترة زمنية محددة. يشير مقام معدل الوفاة إلى عدد أفراد المجتمع الواقع تحت خطر الوفاة بينما يشير البسط إلى عدد تلك الوفيات التي حدثت في المجتمع المشار إليه.

1- معدل الوفاة الخام السنوي : Annual Crude death rate

$$\text{ويساوي} \quad \frac{\text{عدد الوفيات خلال السنة كاملة}}{\text{عدد أفراد المجتمع في منتصف السنة}} \times K$$

حيث K = 1000 عادة.

2- معدل الوفاة المحدد بالعمر Age-Specific Death Rate

$$\text{ويساوي} \quad \frac{\text{عدد الوفيات في السنة من فئة عمرية معينة}}{\text{عدد أفراد المجتمع من الفئة العمرية المعينة في منتصف السنة}} \times K$$

ويمكن تعريف معدلات الوفاة مماثلة للتعريف (2)ولكن لفئات محددة بالجنس أو بالعمر والجنس أو بتصنيفات فئوية معينة مثل.

(3) معدل الوفاة السنوي المحدد بصفة معينة ويساوي

$$\frac{\text{عدد الوفيات في السنة من فئة محددة}}{\text{عدد أفراد المجتمع في الفئة المحددة في منتصف السنة}} \times 1000$$

مثال (4) :

كان عدد سكان مدينة في منتصف عام (1998) 24780 نسمة وبلغ عدد الوفيات فيها ذلك العام 420 وكان عدد الأطفال دون سن السادسة في تلك المدينة 5500 وبلغ عدد الوفيات من الأطفال دون سن

السادسة في ذلك العام 80 فما هو معدل الوفاة الخام السنوي وما هو معدل الوفاة المحدد بعمر دون ست سنوات.

الحل : معدل الوفاة الخام السنوي يساوي :

$$\frac{420}{24780} \times 1000 = 16.95$$

لكل 1000 نسمة.

معدل الوفاة المحدد بعمر دون ست سنوات يساوي :

$$\frac{80}{5500} \times 1000 = 14.5$$

مثال (5) : كان عدد سكان الأردن عام 1990 يساوي 3453000 وكان عدد الوفيات 31080 وكان عدد السكان عام 1989 يساوي 3111000 وبلغ عدد الوفيات ذلك العام 31112 ما معدل الوفاة الخام في كل من العامين ؟

الحل :

معدل الوفاة الخام عام 1990 يساوي :

$$\frac{31080}{3453000} \times 1000 = 9$$

لكل 1000 نسمة.

معدل الوفاة الخام عام 1989 يساوي :

$$\frac{31112}{3111000} \times 1000 = 10$$

لكل 1000 نسمة.

4- معدل الوفاة المعياري Standardized death rate

إنه من الخطأ مقارنة معدلات الوفاة الخام لمجتمعات مختلفة ما لم تكن ظروف هذه المجتمعات الصحية والاجتماعية وغيرها من الظروف مثل فئات الأعمار متماثلة، ولكن من الممكن أن تكون المقارنة ذات معنى عن طريق مقارنة ما يسمى بالمعدلات المعيارية. وهي عبارة عن مقياس واحد يقيس القوى

المؤثرة في الوفيات في مجتمع ما مع تثبيت واحد أو أكثر من عوامل تركيب المجتمع كالعمر أو الجنس أو العِرق. وعادة ما يحسب معدل الوفاة المعياري باستخدام الطريقة المباشرة للتعديل التي تتطلب حساب المعدلات المحددة (بالعمر أو غيره) في المجتمع تحت الدراسة ومن ثم تطبيق هذه المعدلات على المجتمع المعياري (Standard Population).

ومن الأعداد المتوقعة الناتجة من تطبيق المعدلات المحددة على الفئات العمرية في المجتمع المعياري نحسب المعدل الكلي الذي يعطينا قيمة المعدل للمجتمع تحت الدراسة فيما لو كان تركيبه نفس تركيب المجتمع المعياري.

إذا حسـبنا معـدل الوفـاة المعيـاري لكـل مـن مجتمعـين أو أكثـر كـان بالإمكان مقارنـة هذه المعدلات بشكل مباشر.

تختلف الآراء حول المجتمع الذي يمكن اعتباره معياريا. يمكننا في المملكة الأردنية اعتماد التعداد السكاني لسنة 1990 أو لسنة 1995 بعد تعديله لجعله معياريا من مليون كمجتمع معياري ونسميه المليون المعياري.

ولشرح الطريقة المباشرة لحساب معدل الوفاة المعياري نحل المثال التالي :

مثال (6) : تعطي الأعمدة (1)، (2)، (3) الجدول التالي (1) تركيب عدد السكان في احدى محافظات المملكة الأردنية الهاشمية عام 1990 وعدد الوفيات لكل فئة عمرية. احسب معدل الوفاة المعياري لتلك المحافظة.

الحل : نحتاج إلى تعداد السكان في الأردن لعام 1990 ونحسب منه المليون المعياري ، ويظهر هذا في الجدول (2)، حيث حسبنا مدخلات العمود الثالث في هذا الجدول عن طريق :

العدد المعياري للفئة العمرية 4-0 يساوي :

$$\frac{512600}{3453000} \times 1000000 = 148451$$

وهكذا للفئات العمرية الأخرى.

نرتب الحل باكمال الجدول (1) الذي نبنيه كما يلي :

(1) نضع الأفراد حسب الفئات العمرية، العمود (1) ، (2).

(2) نضع عدد الوفيات لكل فئة عمرية (3).

(3) نحسب معدل الوفاة المحدد بالعمر لكل 1000 نسمة العمود (4).

الجدول (1)

(1) الفئة العمرية	(2) عدد الأفراد	(3) عدد الوفيات	(4) معدل الوفاة المحدد بالعمر (1000)	(5) المجتمع المعياري على أساس سكان الأردن عام 1970	(6) عدد الوفيات المتوقع في المجتمع المعياري
0-4	108706	2707	29.5	148451	3696
5-9	108906	654	6.0	144049	864
10-14	125230	250	2.0	1444107	288
15-19	108806	254	2.3	130582	300
20-24	100244	101	1.0	108601	109
25-29	83230	100	1.2	76571	92
30-34	50272	230	4.6	49406	227
35-39	25186	120	4.8	36490	175
40-44	24668	140	5.7	33970	194
45-49	25986	205	8.1	32030	259
50-54	24886	235	9.4	30003	282
55-59	16724	215	12.8	20996	269
60-64	11707	470	40.1	16565	664
65-69	8362	516	61.7	9325	575
70-74	5057	360	71.2	7472	532
75-79	3345	301	90.0	4518	407
80+	5010	460	91.8	6574	603

	835615	7318		1000000	9535

الجدول (2)

الفئة العمرية	عدد السكان [1]	المجتمع المعياري [2]
0-4	512600	148451
5-9	497400	144049
10-14	497600	144107
15-19	450900	130582
20-24	375000	108601
25-29	264400	76571
30-34	170600	49406
35-39	126000	36490
40-44	117300	33970
45-49	110600	32030
50-54	103600	30003
55-59	72500	20996

(1) المصدر: الأمانة العامة للجنة الوطنية للسكان، منشور في النشرة السكانية ، العدد الأول السنة الثانية 1991م ، 1411هـ.
(2) حسبها المؤلف.

60-64	57200	16565
65-69	32200	9325
70-74	25800	7472
75-79	15600	4518
80+	22700	6574
المجموع	3453000	1000000

فمثلا ، أول عدد في العمود (4) هو معدل الوفاة المحدد بالعمر (4-0) ويساوي :

$\frac{2707}{108706} \times 1000 = 29.5$ وهكذا جميع مدخلات العمود (4).

(4) نكتب المجتمع المعياري من المليون المعياري المعطى لنا في السجلات الرسمية حسب الفئات العمرية، ونضعه في العمود (5).

(5) نحسب عدد الوفيات المتوقع في المجتمع المعياري حسب الفئات العمرية ونضعه في العمود (6) ، فمثلا العدد الأول في العمود (6) هو :

$$148451 \times \frac{29.5}{1000} = 3696$$

(6) نجمع مفردات العمود (6) فنحصل على مجموع الوفيات المتوقعة في المجتمع المعياري.

(7) نحسب معدل الوفاة المعياري (المعدل حسب العمر) وتعريفه هو :

$$\frac{\text{عدد الوفيات المتوقعة}}{\text{مجموع أفراد المجتمع المعياري}} \times 1000$$

وهو في مثالنا :

$$\frac{9535}{1000000} \times 1000 = 9.54$$

(8) للمقارنة نحسب معدل الوفاة الخام فنجده :

$$\frac{7318}{835615} \times 1000 = 8.76$$

أي أن معدل الوفاة الخام لسكان المحافظة في مثالنا أصغر من معدل الوفاة المعياري. إن أحد دلالات ذلك أن سكان هذه المحافظة عام 1990 كان فيه نسبة كبار السن أقل من نسبة كبار السن في المجتمع المعياري أي أن سكان هذه المحافظة كانوا أصغر سنا من السكان في المجتمع المعياري، فبينما نجد أنه في المجتمع المعياري يوجد 27.89 لكل 1000 نسمة هم في عمر 65 وأعلى.

$$\frac{9535 + 7472 + 4518 + 6574}{1000000} \times 1000 = 27.89$$

نجد في مثالنا فقط 26.06 لكل 1000 نسمة هم في عمر 65 وأعلى :

$$\frac{8362 + 5057 + 3345 + 5010}{835615} \times 1000 \approx 26.06$$

لاحظ أن الأعمدة (1) ، (2) ، (3) عادة تعطى في السؤال أي تؤخذ من السجلات الرسمية والعمود (5) عادة يحسب من جدول تركيب السكان حسب العمر الذي يؤخذ من السجلات الرسمية في الدولة.

ويبقى فقط حساب العمودين (4)، (6) ثم حساب معدل الوفاة المعياري.

5- معدل وفيات الأمومة : Maternal Mortality

يعرف هذا المعدل بالمعادلة :

$$\frac{\text{عدد وفيات الأمهات لأسباب النفاس خلال عام}}{\text{عدد المواليد الأحياء الذين ولدوا خلال ذلك العام}} \times 1000$$

إن الوفاة لأسباب النفاس أي لأسباب تتعلق بالحمل والولادة تتضمن الوفيات التي تحدث عند الولادة أو نتيجة مضاعفات قد تنشأ في وقت لاحق أو وقت سابق لها بسبب الحمل أو غيره، وقد حددت منظمة الصحة العالمية WHO حالات الوفاة لأسباب النفاس بأنها تلك الوفاة التي تحدث للمرأة الحامل ضمن 42 يوما من انتهاء الحمل.

هناك تحديدات وملحوظات حول هذا المعدل :

(1) حالات الاسقاط وولادة المولود غير الحي ليست محسوبة في المقام ولذلك يصغر المقام وينتج معدل أعلى من الحقيقة.

(2) وفاة المرأة لأسباب الحمل والنفاس تحسب مرة واحدة في كل حالة، مع أن الولادة لمواليد الأحياء ممكن أن تعطي توائم أو ثلاثة مواليد أو أربعة، وبالتالي يكبر المقام ويصغر المعدل.

(3) كثيرا من الأحيان لا يسجل الأهل الولادة الحية (وخاصة في المجتمعات البدائية) مما يؤدي إلى تصغير المقام.

(4) من الممكن أن تحدث الوفاة في سنة غير السنة التي حدثت فيها الولادة ولذلك يصغر البسط في هذه الحالة.

باعتبار الملاحظات السابقة في مجملها فإن النقص والزيادة في البسط أو المقام ستعادل بعضها البعض ونحصل على معدل قريب من المعدل الحقيقي.

6- معدل وفيات الأطفال الرضع : Infant Mortality Rate

$$\text{ويساوي} \quad \frac{\text{عدد الوفيات في الأطفال الذين تقل أعمارهم عن سنة واحدة خلال عام}}{\text{عدد الأطفال المولودين أحياء في نفس العام}} \times 1000$$

إن استعمال هذا المعدل وتفسيره يخضع للمحددات التي ذكرناها بعد تعريف معدل وفيات الأمومة . ومما يستدعي الانتباه أن الأطفال المولودين أحياء في سنة معينة ربما يموت بعضهم في السنة التالية وقبل أن يبلغوا من العمر عاما واحدا وكثير من الأطفال الرضع الذين يموتون في سنة معينة يكونون قد ولدوا في العام السابق.

إن مثل هذا التحديد لا يؤثر كثيرا في المجتمعات التي معدل الولادة فيها ثابت تقريبا.

مثال (7) : إذا كان عدد الأطفال المولودين أحياء في الأردن عام 1990 يساوي 113911 وبلغ عدد الوفيات من بينهم 5127، فما هو معدل وفيات الأطفال الرضع في عام 1990 :

الحل : معدل وفيات الأطفال الرضع يساوي :

$$\frac{5127}{113911} \times 1000 = 45 \text{ لكل } 1000$$

7- معدل وفيات الأطفال حديثي الولادة : Neonatal Mortality Rate

عدد الوفيات في الأطفال في سنة معينة الذين تقل أعمارهم عن 28 يوما

ويساوي ـــــــــــــــــــــــــــــــــ ×1000

عدد الأطفال المولودين أحياء في نفس السنة

8- معدل وفيات الاسقاط Fetal death rate

عدد حالات الاسقاط خلال عام

ويساوي ـــــــــــــــــــــــ × 1000

عدد حالات الولادة خلال ذلك العام

ملاحظة : (1) إن المقام هو عبارة عن : عدد حالات الاسقاط + عدد الأطفال المولودين أحياء.

(2) عادة ما نضيف وصفا لحالة الإسقاط كأن نقول :حالات الإسقاط (ذات وزن محدد أو أكثر) مثل : حالات الإسقاط بوزن 600غم أو أكثر.

(3) هناك فروق بين بعض الأقطار في تعريف حالة إسقاط الجنين تعتمد على فترة الحمل ولكن الأكثر شيوعا هو أية حالة إسقاط بغض النظر عن فترة الحمل.

9- نسبة وفيات الإسقاط Fetal death ratio

عدد حالات الإسقاط خلال عام

وتساوي ـــــــــــــــــــــــ × 1000

عدد الأطفال الذين ولدوا أحياء في ذلك العام

أي أننا لا نضيف إلى المقام عدد حالات الإسقاط.

10- نسبة الوفاة بسبب محدد Cause - of - Death ratio

عدد الوفيات التي تعزى لمرض معين خلال عام

ويساوي ———————— × 100

عدد جميع الوفيات في ذلك العام

وتحسب هذه النسبة لمعرفة أهمية أو شدة خطورة مرض معين في تسببه لحالة وفاة بالنسبة للوفيات جميعها مهما اختلفت حالاتها.

11- نسبة الوفاة التناسبية Proportional Mortality Rate

عدد الوفيات من عمر 50 سنة أو أكثر

وتساوي ———————— × 100

عدد جميع الوفيات

وتستعمل هذه النسبة كمقياس منفرد لمقارنة الحالة الصحية الكلية بين المجتمعات المختلفة.

مثال (8) : كان عدد سكان الأردن من عمر 50 سنة فأكثر عام 1990 يساوي 329600

وبلغت الوفيات تلك السنة من الفئة العمرية هذه 6261 ومجموع الوفيات الكلي ذلك العام 24861فما نسبة الوفاة التناسبية ؟

وما معدل الوفاة المحدد بعمر 50 سنة فأكثر ؟

الحل : نسبة الوفاة التناسبية تساوي :

عدد الوفيات من عمر 50 سنة فأكثر

———————— × 100

عدد جميع الوفيات

ويساوي : $\frac{6261}{24861} \times 100$ ويساوي 25.2%

معدل الوفاة المحدد بعمر 50 سنة فأكثر يساوي :

$$\frac{6261}{329600} \times 1000 = 19$$

لكل 1000 نسمة.

3-14 إحصاءات الخصوبة Fertility Statistics

إن معرفة الخصوبة ونسبة المواليد في مجتمع ما أمر هام للعاملين في الحقل الصحي وذلك في التخطيط لإنشاء الخدمات ومراكز الرعاية الصحية وتوفير المرافق الضرورية للأمهات والأطفال الرضع والأطفال.

1- معدل الولادة الخام Crude Birth Rate

$$\text{ويساوي} = \frac{\text{عدد المواليد الأحياء خلال عام}}{\text{عدد السكان الكلي في منتصف العام}} \times 1000$$

2- معدل الخصوبة العام General Fertility Rate

$$\text{ويساوي} = \frac{\text{عدد المواليد الأحياء خلال عام}}{\text{عدد النساء في سن الحمل في منتصف العام}} \times 1000$$

ويعرف سن الحمل ما بين عمر 15 إلى 44 أو عمر 15 إلى 49 سنة.

3- معدل الخصوبة المحدد بالعمر Age-Specific Fertility

بما أن معدل الحمل ليس متجانسا على مدى سن الحمل (بمعنى أن معدل الحمل للنساء في سن 24 سنة أكثر من معدل الحمل للنساء فوق سن 35 سنة) فقد اصبح من المرغوب فيه أن نحسب معدلا يسمح بتحليل معدلات الخصوبة لفئات عمرية مختلفة، أي أننا نحتاج لحساب معدل خصوبة محددا بالعمر، وتعريفه :

$$\text{معدل الخصوبة المحدد بالعمر} = \frac{\text{عدد المواليد الأحياء من نساء في عمر معين خلال عام}}{\text{عدد جميع النساء في ذلك العمر عند منتصف العام}} \times 100$$

مثال (9) : بلغ عدد سكان الأردن عام 1995 ما مجموعة 4098000 وبلغ عدد المواليد الأحياء في ذلك العام 141784 فما هو معدل الولادة الخام ؟

الحل : معدل الولادة الخام لعام 1995 يساوي :

$$\frac{141784 \times 1000}{4098000} = 34.6$$

لكل 1000 نسمة.

مثال (10) : كان عدد الإناث في عمر 15-49 سنة في منتصف 1995 يساوي 24.1% من عدد السكان وكان عدد المواليد الأحياء ذلك العام 141784 أوجد معدل الخصوبة العام.

الحل : عدد الإناث في عمر 15-49 سنة يساوي $4098000 \times \frac{24.1}{100} = 987618$

معدل الخصوبة العام يساوي :

$$\frac{141784 \times 1000}{987618} = 144$$

لكل 1000 امرأة.

مثال (11) : كان عدد الإناث في عمر (25-29) سنة في منتصف عام 1995 يساوي 182300 وكان عدد المواليد الأحياء من هؤلاء النسوة 27300 مولودا.

أوجد معدل الخصوبة المحدد بعمر (25-29) سنة.

الحل :

$$\frac{27300}{182300} \times 1000 = 149.7$$

لكل 1000 امرأة في عمر (25-29).

مثال (12) : كان عدد الإناث في الفئة العمرية (40-44) في منتصف عام 1990 يساوي 60300 وكان عدد المواليد الأحياء من هؤلاء النسوة يساوي 280 مولودا.

أوجد معدل الخصوبة المحدد بعمر (40-44).

الحل :

$$\frac{280}{6300} \times 1000 = 4.6$$

لكل 1000 امرأة في هذه الفئة العمرية.

ويمكن حساب المعدلات المحددة بالعمر لسنوات مفردة من العمر (مثلا عمر 22 سنة) أو فترة عمرية، وأكثر هذه المعدلات تلك المحسوبة للمجموعات ضمن فترة خمس سنوات. ويمكن حساب معدل الخصوبة المحدد بالعرق أو الحالة الاقتصادية الاجتماعية أو سمات ديموغرافية أخرى متعددة.

4-14 : إحصاءات الأمراض Morbidity Statistics

تعتبر هذه الإحصاءات من المواضيع التي تهم العاملين في المجال الصحي والذين يهتمون في تحليل الوضع الصحي في المجتمع، وبوجه عام فإن البيانات عن الأمراض في مجتمع ما لا تكون متوفرة ومتكاملة كما هو الحال في بيانات الولادات والوفيات وذلك بسبب عدم استكمال الإخبار عن المرض وتفاوت التشريعات بوجوب الإخبار عن المرض. وأهم المعدلات المستعملة في دراسة الأمراض في مجتمع معين هما معدل الإصابات ومعدل الانتشار.

1- معدل الإصابات Incidence Rate

$$\text{ويساوي} \quad \frac{\text{معدل الإصابات الجديدة من مرض معين خلال عام}}{\text{عدد السكان في منتصف العام}} \times 100$$

إن هذا المعدل يقيس درجة حدوث حالات جديدة من الإصابة بالمرض وبالتالي فهو يساعد في تحديد الحاجة لاتخاذ إجراءات احترازية. إنه مقياس مفيد في حالات الأمراض المزمنة والأمراض الحادة.

2- معدل الانتشار Prevalence Rate

يقيس هذا المعدل مقدار انتشار مرض معين في بلد ما وهو يساوي :

$$\frac{\text{عدد الإصابات الموجودة (قديمة أو جديدة) في نقطة زمنية معينة}}{\text{عدد السكان في تلك النقطة الزمنية}} \times 1000$$

ويستعمل هذا المعدل بصفة خاصة في دراسة الأمراض المزمنة مع أنه يمكن استعماله أيضا في الأمراض الحادة.

3- نسبة حالات الهلاك Case-Fatality Ratio

وهذه النسبة مفيدة في التعرف على مدى نجاح برنامج مكافحة مرض معين وتعرف هذه النسبة بـ :

عدد حالات الوفاة بسبب مرض معين

$$\frac{\text{عدد حالات الوفاة بسبب مرض معين}}{\text{عدد حالات الإصابة بهذا المرض}} \times 1000$$

عدد حالات الإصابة بهذا المرض

والفترة الزمنية التي تقاس بها هذه النسبة اختيارية وتعتمد على طبيعة المرض ويمكن أن تكون لعدة سنوات في حالة المرض المستوطن. ومن الملاحظ أن هذه النسبة يمكن تفسيرها على أنها احتمال الوفاة بعد الإصابة بالمرض المعني، ولذلك فهي تكشف مدى خطورة ذلك المرضى.

4- نسبة عدم النضوح (خداج) : Immaturity Ratio

وتساوي

$$\frac{\text{عدد الأطفال المولودين أحياء بوزن أقل من 2500 غم خلال عام}}{\text{العدد الكلي للأطفال المولودين أحياء خلال ذلك العام}} \times 1000$$

مثال (13) :

بلغ عدد الأطفال الذين أصيبوا بالسرطان في الأردن عام 1995 مائتي طفل، ما معدل الإصابة بهذا المرض بين الأطفال.

الحل : معدل الإصابات بالسرطان بين الأطفال =

$$\frac{\text{معدل الإصابات الجديدة بالسرطان عام 1995}}{\text{عدد السكان في منتصف العام}} \times 1000$$

$$\frac{200}{4098000} \times 1000 \approx 0.05$$ لكل ألف

لاحظ أن عدد سكان الأردن منتصف عام 1995 هو 4098000 كما ورد في الأمثلة السابقة وحسب النشرة السكانية في الأردن.

مثال (14) : بينت إحصائية عن حالات الإصابة بمرض الإيدز في احدى المدن أن عدد المصابين بذلك المرض وقت التعداد هو 3417 حالة وكان عدد سكان المدينة 278518 فما معدل انتشار ذلك المرض.

الحل : معدل الانتشار = $1000 \times \frac{3417}{278518}$

= 12.4 إصابة لكل 1000.

مثال (15) : بلغ عدد الأطفال المولودين أحياء في عام 1998 في مدينة 5324 وكان من بينهم 1476 طفلا بوزن أقل من 2500 غم، فإذا علم أن عدد سكان المدينة في منتصف عام 1998 هو 133600.

أ- أوجد معدل الولادة الخام.

ب- أوجد نسبة عدم النضوج.

الحل :

أ- معدل الولادة الخام = $1000 \times \frac{5324}{133600}$

= 39.8 لكل 1000

ب- نسبة عدم النضوج = $1000 \times \frac{1476}{5324}$

= 277.2 لكل 1000.

مثال (16) : في مثال (14) إذا توفى بمرض الإيدز في تلك المدينة 907 فما نسبة حالات الهلاك ؟

الحل : نسبة حالات الهلاك = $1000 \times \frac{907}{3417}$

= 265.4 لكل 1000.

تمارين

1-14 : يمثل الجدول التالي تقدير توزيع السكان في الأردن حسب العمر والجنس لعام 1995 -

المصدر : الأمانة العامة للجنة الوطنية للسكان. النشرة السكانية - العدد الأول - السنة الثانية 1991م ، 1411هـ.

الفئة العمرية	الذكور	الإناث	المجموع
المجموع	2102900	1994800	4097800
0-4	326800	311100	637900
5-9	264100	257500	521600
10-14	262600	244900	507500
15-19	265300	239800	505100
20-24	238500	220800	459300
25-29	200900	182300	383200
30-34	141600	130400	271900
35-39	84000	91300	175200
40-44	63800	64800	128600
45-49	57300	60200	117500
50-54	55600	52200	108900
55-59	50900	48900	99800
60-64	34300	33300	67700
65-69	26300	24100	50500
70-74	14900	12000	28000
75-79	8400	9800	18200

80+	7600	9300	16900

أجب عن الأسئلة الآتية المتعلقة بالجدول والمعطيات على فرض أن عدد السكان مقدر في 1 تموز 1995.

(1) إذا كان عدد الوفيات 22538 فما هو معدل الوفاة الخام ؟

(2) إذا كان عدد الوفيات من عمر 79-75 سنة يساوي 1458، فما هو معدل الوفاة المحدد بالعمر (79-75) ؟

(4) إذا كان عدد المواليد الأحياء من النساء في عمر (24-20) يساوي 33562 فما معدل الخصوبة المحدد بالعمر (24-20) ؟

(5) إذا كان عدد المواليد الأحياء من النساء في عمر (44-40) يساوي 273 فما هو معدل الخصوبة المحدد بالعمر (44-40) ؟

(6) إذا كان عدد الوفيات في عام 1995 يساوي 33885 فما هو معدل الوفاة العام.

(7) إذا كان عدد وفيات الإناث في عام 1995 يساوي 16159 فما هو معدل وفيات الإناث العام ؟

(8) ما معدل وفيات الذكور العام لسنة 1995 ؟

2-14 : يعطي الجدول التالي بيانات عن احدى المحافظات لعام 1997.

البيان	العدد
تقدير عدد السكان في 1 تموز (يوليو)	720800
مجموع المواليد أحياء	23900
مجموع المواليد غير الناضجين أقل من 2500 غم (خداج)	1000
مجموع حالات الاسقاط	810
مجموع الوفيات (جميع الأعمار)	7415
الوفيات للأطفال من عمر أقل من سنة	392
الوفيات للأطفال من عمر أقل من 28 يوما	310

عدد الوفيات للمواليد غير الناضجين (خداج)	40
عدد الأمهات المتوفيات بالنفاس	76
عدد الوفيات بالسرطان	1516

من هذه البيانات احسب ما يلي :

أ- معدل الوفاة الخام.

ب- معدل وفيات الأمومة.

جـ- معدل وفيات الأطفال الرضع.

د- معدل وفيات الأطفال حديثي الولادة.

هـ- معدل وفيات الإسقاط

و- نسبة وفيات السرطان

ز- نسبة وفيات الإسقاط

ح- نسبة عدم النضوج.

3-14 : إضافة إلى البيانات في تمرين (14-2) لديك البيانات عن المحافظة نفسها لعام 1997 :

عدد النساء في عمر (15-49) سنة في منتصف العام	252750
عدد النساء في عمر 22 سنة في منتصف العام	9425
مجموع المواليد الأحياء من نساء في عمر 22 سنة	1289
عدد الإصابات بالسرطان	2437

أوجد :

أ- معدل الولادة الخام.

ب- معدل الخصوبة العام.

جـ- معدل الخصوبة المحدد بعمر 22 سنة.

د- معدل الزيادة الطبيعية والتي تعرف بزيادة معدل المواليد الخام على معدل الوفيات الخام.

هـ- معدل الإصابات بالسرطان.

4-14 : يمثل الجدول التالي توزيع السكان في إحدى محافظات المملكة الأردنية الهاشمية حسب العمر والجنس لعام 1995 وكذلك الوفيات حسب الفئات العمرية.

(1) الفئة العمرية	(2) عدد السكان	(3) عدد الوفيات
المجموع	1717246	14038
0-4	280676	5785
5-9	242288	969
10-14	203000	406
15-19	207091	435
20-24	183720	184
25-29	160944	176
30-34	106041	445
35-39	68328	258
40-44	51440	230
45-49	50525	375
50-54	46283	368
55-59	44910	536
60-64	27080	945
65-69	20200	1020
70-74	11200	762
75-79	7098	568
80+	6422	576

أ- أوجد المجتمع المعياري (المليون المعياري) على أساس سكان الأردن حسب العمر لعام 1995 المعطى في تمرين (14-1) وضعه في عمود (5) على الجدول المعطى.

ب- أوجد معدل الوفاة المحدد بالعمر وضعه في عمود (4).

جـ- أوجد عدد الوفيات المتوقع في المجتمع المعياري وضعه في العمود (6).

د- أوجد معدل الوفاة العام.

هـ- أوجد معدل الوفاة المعياري.

الملاحق
الجداول الإحصائية

الجدول (I) : جدول الأعداد العشوائية

الجدول (II) : احتمالات ذات الحدين.

الجدول (III) : احتمالات بواسون التراكمية

الجدول (IV) : المساحة تحت المنحنى الطبيعي المعياري

الجدول (V) : يعطي القيم على المحور الأفقي t وهي : $t\,[\lambda\,;\,n]$ التي يكون إلى يسارها λ من المساحة تحت توزيع t ذي درجات الحرية n.

الجدول (VI) : يعطي القيم على المحور الأفقي χ^2 وهي $\chi^2\,[\lambda\,;\,n]$ التي يكون إلى يسارها λ من المساحة تحت منحنى توزيع χ^2 ذي درجات الحرية n.

الجدول (VII): يعطي القيم على المحور الأفقي F أي $F\,[\lambda;\,n_1,\,n_2]$ التي يكون إلى يسارها λ من المساحة تحت منحنى توزيع F ذي n_1 درجات حرية في البسط، n_2 درجات حرية في المقام.

الجدول (VIII) : القيم الطبيعية Normal scores.

الجدول (IX) : تحويل r إلى فيشر z . $Z = \frac{1}{2}\ln\frac{1+r}{1-r}$

الجدول (I) : جدول الأعداد العشوائية

53 74 23 99 67	61 32 28 69 84	94 62 67 86 24	98 33 41 19 95	47 53 53 38 09
63 38 06 86 54	99 00 65 26 94	02 82 90 23 07	79 62 67 80 60	75 91 12 81 19
35 30 58 21 46	06 72 17 10 94	25 21 31 75 96	49 28 24 00 49	55 65 79 78 07
63 43 36 82 69	65 51 18 37 88	61 38 44 12 45	32 92 85 88 65	54 34 81 85 35
98 25 37 55 26	01 91 82 81 46	74 71 12 94 97	24 02 71 37 07	03 92 18 66 75
02 63 21 17 69	71 50 80 89 56	38 15 70 11 48	43 40 45 86 98	00 83 26 91 03
64 55 22 21 82	48 22 28 06 00	61 54 13 43 91	82 78 12 23 29	06 66 24 12 27
85 07 26 13 89	01 10 07 82 04	59 63 69 36 03	69 11 15 83 80	13 29 54 19 28
58 54 16 24 15	51 54 44 82 00	62 61 65 04 69	38 18 65 18 97	85 72 13 49 21
34 85 27 84 87	61 48 64 56 26	90 18 48 13 26	37 70 15 24 57	65 65 80 39 07
03 92 18 27 46	57 99 16 96 56	30 33 72 85 22	84 64 38 56 98	99 01 30 98 64
62 95 30 27 59	37 75 41 66 48	86 97 80 61 45	23 53 04 01 63	45 76 08 64 27
08 45 93 15 22	60 21 75 46 91	98 77 27 85 42	28 88 61 08 84	69 62 03 42 73
07 08 55 18 40	45 44 75 13 90	24 94 96 61 02	57 55 66 83 15	73 42 37 11 61
01 85 89 95 66	51 10 19 34 88	15 84 97 19 75	12 76 39 43 78	65 63 91 08 25
72 84 71 14 35	19 11 58 49 26	50 11 17 17 76	86 31 57 20 18	95 60 78 46 75
88 78 28 16 84	13 52 53 94 53	75 45 69 30 96	73 89 65 70 31	99 17 43 48 76
45 17 75 65 57	28 43 19 72 12	25 12 74 75 67	60 40 60 81 19	24 62 01 61 16
96 76 28 12 54	22 01 11 94 25	71 96 16 16 88	68 64 36 74 45	19 59 50 88 92
43 31 67 72 30	24 02 94 08 63	38 32 36 66 02	69 36 38 25 39	48 03 45 15 22
50 44 66 44 21	66 06 58 05 62	68 15 54 35 02	42 35 48 96 32	14 52 41 52 48
22 66 22 15 86	26 63 74 41 99	58 42 36 72 24	48 37 52 18 51	03 37 18 39 11
96 24 40 14 51	23 22 30 88 57	95 67 47 29 83	94 69 40 06 07	18 16 36 78 86
31 73 91 61 19	60 20 72 93 48	98 57 07 34 69	65 95 39 69 58	56 80 30 19 44
78 60 73 99 84	43 89 94 36 45	56 69 47 07 41	90 22 91 07 12	78 35 34 08 72
84 37 90 61 56	70 10 23 98 05	85 11 34 76 60	76 48 45 34 60	01 64 18 39 96
36 67 10 08 23	98 93 35 08 86	99 29 76 29 81	33 34 91 58 93	63 14 52 32 52
07 28 59 07 48	89 64 58 89 75	83 85 62 27 89	30 14 78 56 27	86 63 59 80 02
10 15 83 87 60	79 24 31 66 56	21 48 24 06 93	91 98 94 05 49	01 47 59 38 00
55 19 68 97 65	03 73 52 16 56	00 53 55 90 27	33 42 29 38 87	22 13 88 83 34
53 81 29 13 39	35 01 20 71 34	62 33 74 82 14	43 73 19 09 03	56 54 29 56 93
51 86 32 68 92	33 98 74 66 99	40 14 71 94 58	45 94 19 38 81	14 44 99 81 07
35 91 70 29 13	80 03 54 07 27	96 94 78 32 66	50 95 52 74 33	13 80 55 62 54
37 71 67 95 13	20 02 44 95 94	64 85 04 05 72	01 32 90 76 14	53 89 74 60 41
93 66 13 83 27	92 79 64 64 72	28 54 96 53 84	48 14 52 98 94	56 07 93 89 30
02 96 08 45 65	13 05 00 41 84	93 07 54 72 59	21 45 57 09 77	19 48 56 27 44
49 83 43 48 35	82 88 33 69 96	72 36 04 19 76	47 45 15 18 60	82 11 08 95 97
84 60 71 62 46	40 80 81 30 37	34 39 23 05 38	25 15 35 71 30	88 12 57 21 77
18 17 30 88 71	44 91 14 88 47	89 23 30 63 15	56 34 20 47 89	99 82 93 24 98
79 69 10 61 78	71 32 76 95 62	87 00 22 58 40	92 54 01 74 25	43 11 71 99 31
75 93 36 57 83	56 20 14 82 11	74 21 97 90 65	96 42 68 63 86	74 54 13 26 94
38 30 92 29 03	06 28 81 39 38	62 25 06 84 63	61 29 08 93 67	04 32 92 08 09
51 29 50 10 34	31 57 75 95 80	51 97 02 74 77	76 15 48 49 44	18 55 63 77 09
21 31 38 86 24	37 79 81 53 74	73 24 16 10 33	52 83 90 94 76	70 47 14 54 36
29 01 23 87 88	58 02 39 37 67	42 10 14 20 92	16 55 23 42 45	54 96 09 11 06
95 33 95 22 00	18 74 72 00 18	38 79 58 69 32	81 76 80 26 92	82 80 84 25 39
90 84 60 79 80	24 36 59 87 38	82 07 53 89 35	96 35 23 79 18	05 98 90 07 35
46 40 62 98 82	54 97 20 56 95	15 74 80 08 32	16 46 70 50 80	67 72 16 42 79
20 31 89 03 43	38 46 82 68 72	32 14 82 99 70	80 60 47 18 97	63 49 30 21 30
71 59 73 05 50	08 22 23 71 77	91 01 93 20 49	82 96 59 26 94	66 39 67 98 60

الجدول (II) : احتمالات ذات الحدين.

n	x	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.75	0.8	0.9
1	0	0.900	0.800	0.750	0.700	0.600	0.500	0.400	0.300	0.250	0.200	0.100
	1	0.100	0.200	0.250	0.300	0.400	0.500	0.600	0.700	0.750	0.800	0.900
2	0	0.810	0.640	0.563	0.490	0.360	0.250	0.160	0.090	0.063	0.040	0.010
	1	0.180	0.320	0.375	0.420	0.480	0.500	0.480	0.420	0.375	0.320	0.180
	2	0.010	0.040	0.063	0.090	0.160	0.250	0.360	0.490	0.563	0.640	0.810
3	0	0.729	0.512	0.422	0.343	0.216	0.125	0.064	0.027	0.016	0.008	0.001
	1	0.243	0.384	0.422	0.441	0.432	0.375	0.288	0.189	0.141	0.096	0.027
	2	0.027	0.096	0.141	0.189	0.288	0.375	0.432	0.441	0.422	0.384	0.243
	3	0.001	0.008	0.016	0.027	0.064	0.125	0.216	0.343	0.422	0.512	0.729
4	0	0.656	0.410	0.316	0.240	0.130	0.063	0.026	0.008	0.004	0.002	0.000
	1	0.292	0.410	0.422	0.412	0.346	0.250	0.154	0.076	0.047	0.026	0.004
	2	0.049	0.154	0.211	0.265	0.346	0.375	0.346	0.265	0.211	0.154	0.049
	3	0.004	0.026	0.047	0.076	0.154	0.250	0.346	0.412	0.422	0.410	0.292
	4	0.000	0.002	0.004	0.008	0.026	0.063	0.130	0.240	0.316	0.410	0.656
5	0	0.590	0.328	0.237	0.168	0.078	0.031	0.010	0.002	0.001	0.000	0.000
	1	0.328	0.410	0.396	0.360	0.259	0.156	0.077	0.028	0.015	0.006	0.000
	2	0.073	0.205	0.264	0.309	0.346	0.312	0.230	0.132	0.088	0.051	0.008
	3	0.008	0.051	0.088	0.132	0.230	0.312	0.346	0.309	0.264	0.205	0.073
	4	0.000	0.006	0.015	0.028	0.077	0.156	0.259	0.360	0.396	0.410	0.328
	5	0.000	0.000	0.001	0.002	0.010	0.031	0.078	0.168	0.237	0.328	0.590
6	0	0.531	0.262	0.178	0.118	0.047	0.016	0.004	0.001	0.000	0.000	0.000
	1	0.354	0.393	0.356	0.303	0.187	0.094	0.037	0.010	0.004	0.002	0.000
	2	0.098	0.246	0.297	0.324	0.311	0.234	0.138	0.060	0.033	0.015	0.001
	3	0.015	0.082	0.132	0.185	0.276	0.313	0.276	0.185	0.132	0.082	0.015
	4	0.001	0.015	0.033	0.060	0.138	0.234	0.311	0.324	0.297	0.246	0.098
	5	0.000	0.002	0.004	0.010	0.037	0.094	0.187	0.303	0.356	0.393	0.354
	6	0.000	0.000	0.000	0.001	0.004	0.016	0.047	0.118	0.178	0.262	0.531
7	0	0.478	0.210	0.133	0.082	0.028	0.008	0.002	0.000	0.000	0.000	0.000
	1	0.372	0.367	0.311	0.247	0.131	0.055	0.017	0.004	0.001	0.000	0.000
	2	0.124	0.275	0.311	0.318	0.261	0.164	0.077	0.025	0.012	0.004	0.000
	3	0.023	0.115	0.173	0.227	0.290	0.273	0.194	0.097	0.058	0.029	0.003
	4	0.003	0.029	0.058	0.097	0.194	0.273	0.290	0.227	0.173	0.115	0.023
	5	0.000	0.004	0.012	0.025	0.077	0.164	0.261	0.318	0.311	0.275	0.124
	6	0.000	0.000	0.001	0.004	0.017	0.055	0.131	0.247	0.311	0.367	0.372
	7	0.000	0.000	0.000	0.000	0.002	0.008	0.028	0.082	0.133	0.210	0.478

تابع الجدول (II)

n	x	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.75	0.8	0.9
8	0	0.430	0.168	0.100	0.058	0.017	0.004	0.001	0.000	0.000	0.000	0.000
	1	0.383	0.336	0.267	0.198	0.090	0.031	0.008	0.001	0.000	0.000	0.000
	2	0.149	0.294	0.311	0.296	0.209	0.109	0.041	0.010	0.004	0.001	0.000
	3	0.033	0.147	0.208	0.254	0.279	0.219	0.124	0.047	0.023	0.009	0.000
	4	0.005	0.046	0.087	0.136	0.232	0.273	0.232	0.136	0.087	0.046	0.005
	5	0.000	0.009	0.023	0.047	0.124	0.219	0.279	0.254	0.208	0.147	0.033
	6	0.000	0.001	0.004	0.010	0.041	0.109	0.209	0.296	0.311	0.294	0.149
	7	0.000	0.000	0.000	0.001	0.008	0.031	0.090	0.198	0.267	0.336	0.383
	8	0.000	0.000	0.000	0.000	0.001	0.004	0.017	0.058	0.100	0.168	0.430
9	0	0.387	0.134	0.075	0.040	0.010	0.002	0.000	0.000	0.000	0.000	0.000
	1	0.387	0.302	0.225	0.156	0.060	0.018	0.004	0.000	0.000	0.000	0.000
	2	0.172	0.302	0.300	0.267	0.161	0.070	0.021	0.004	0.001	0.000	0.000
	3	0.045	0.176	0.234	0.267	0.251	0.164	0.074	0.021	0.009	0.003	0.000
	4	0.007	0.066	0.117	0.172	0.251	0.246	0.167	0.074	0.039	0.017	0.001
	5	0.001	0.017	0.039	0.074	0.167	0.246	0.251	0.172	0.117	0.066	0.007
	6	0.000	0.003	0.009	0.021	0.074	0.164	0.251	0.267	0.234	0.176	0.045
	7	0.000	0.000	0.001	0.004	0.021	0.070	0.161	0.267	0.300	0.302	0.172
	8	0.000	0.000	0.000	0.000	0.004	0.018	0.060	0.156	0.225	0.302	0.387
	9	0.000	0.000	0.000	0.000	0.000	0.002	0.010	0.040	0.075	0.134	0.387
10	0	0.349	0.107	0.056	0.028	0.006	0.001	0.000	0.000	0.000	0.000	0.000
	1	0.387	0.268	0.188	0.121	0.040	0.010	0.002	0.000	0.000	0.000	0.000
	2	0.194	0.302	0.282	0.233	0.121	0.044	0.011	0.001	0.000	0.000	0.000
	3	0.057	0.201	0.250	0.267	0.215	0.117	0.042	0.009	0.003	0.001	0.000
	4	0.011	0.088	0.146	0.200	0.251	0.205	0.111	0.037	0.016	0.006	0.000
	5	0.001	0.026	0.058	0.103	0.201	0.246	0.201	0.103	0.058	0.026	0.001
	6	0.000	0.006	0.016	0.037	0.111	0.205	0.251	0.200	0.146	0.088	0.011
	7	0.000	0.001	0.003	0.009	0.042	0.117	0.215	0.267	0.250	0.201	0.057
	8	0.000	0.000	0.000	0.001	0.011	0.044	0.121	0.233	0.282	0.302	0.194
	9	0.000	0.000	0.000	0.000	0.002	0.010	0.040	0.121	0.188	0.268	0.387
	10	0.000	0.000	0.000	0.000	0.000	0.001	0.006	0.028	0.056	0.107	0.349
11	0	0.314	0.086	0.042	0.020	0.004	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.384	0.236	0.155	0.093	0.027	0.005	0.001	0.000	0.000	0.000	0.000
	2	0.213	0.295	0.258	0.200	0.089	0.027	0.005	0.001	0.000	0.000	0.000
	3	0.071	0.221	0.258	0.257	0.177	0.081	0.023	0.004	0.001	0.000	0.000
	4	0.016	0.111	0.172	0.220	0.236	0.161	0.070	0.017	0.006	0.002	0.000
	5	0.002	0.039	0.080	0.132	0.221	0.226	0.147	0.057	0.027	0.010	0.000
	6	0.000	0.010	0.027	0.057	0.147	0.226	0.221	0.132	0.080	0.039	0.002
	7	0.000	0.002	0.006	0.017	0.070	0.161	0.236	0.220	0.172	0.111	0.016
	8	0.000	0.000	0.001	0.004	0.023	0.081	0.177	0.257	0.258	0.221	0.071
	9	0.000	0.000	0.000	0.001	0.005	0.027	0.089	0.200	0.258	0.295	0.213
	10	0.000	0.000	0.000	0.000	0.001	0.005	0.027	0.093	0.155	0.236	0.384
	11	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.020	0.042	0.086	0.314

الجدول (III) : احتمالات بواسون التراكمية

	μ									
k	.50	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0
0	.607	.368	.135	.050	.018	.007	.002	.001	.000	.000
1	.910	.736	.406	.199	.092	.040	.017	.007	.003	.001
2	.986	.920	.677	.423	.238	.125	.062	.030	.014	.006
3	.998	.981	.857	.647	.433	.265	.151	.082	.042	.021
4	1.000	.996	.947	.815	.629	.440	.285	.173	.100	.055
5	1.000	.999	.983	.961	.785	.616	.446	.301	.191	.116
6	1.000	1.000	.995	.966	.889	.762	.606	.450	.313	.207
7	1.000	1.000	.999	.988	.949	.867	.744	.599	.453	.324
8	1.000	1.000	1.000	.996	.979	.932	.847	.729	.593	.456
9	1.000	1.000	1.000	.999	.992	.968	.916	.830	.717	.587
10	1.000	1.000	1.000	1.000	.997	.986	.957	.901	.816	.706
11	1.000	1.000	1.000	1.000	.999	.995	.980	.947	.888	.803
12	1.000	1.000	1.000	1.000	1.000	.998	.991	.973	.936	.876
13	1.000	1.000	1.000	1.000	1.000	.999	.996	.987	.966	.926
14	1.000	1.000	1.000	1.000	1.000	1.000	.999	.994	.983	.959
15	1.000	1.000	1.000	1.000	1.000	1.000	.999	.998	.992	.978
16	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.996	.989
17	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.998	.995
18	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.998
19	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999
20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

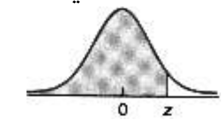
k	10.0	11.0	12.0	13.0	14.0	15.0
			μ			
2	.003	.001	.001	.000	.000	.000
3	.010	.005	.002	.001	.000	.000
4	.029	.015	.008	.004	.002	.001
5	.067	.038	.020	.011	.006	.003
6	.130	.079	.046	.026	.014	.008
7	.220	.143	.090	.054	.032	.018
8	.333	.232	.155	.100	.062	.037
9	.458	.341	.242	.166	.109	.070
10	.583	.460	.347	.252	.176	.118
11	.697	.579	.462	.353	.260	.185
12	.792	.689	.576	.463	.358	.268
13	.864	.781	.682	.573	.464	.363
14	.917	.854	.772	.675	.570	.466
15	.951	.907	.844	.764	.669	.568
16	.973	.944	.899	.835	.756	.664
17	.986	.968	.937	.890	.827	.749
18	.993	.982	.963	.930	.883	.819
19	.997	.991	.979	.957	.923	.875
20	.998	.995	.988	.975	.952	.917
21	.999	.998	.994	.986	.971	.947
22	1.000	.999	.997	.992	.983	.967
23	1.000	1.000	.999	.996	.991	.981
24	1.000	1.000	.999	.998	.995	.989
25	1.000	1.000	1.000	.999	.997	.994
26	1.000	1.000	1.000	1.000	.999	.997
27	1.000	1.000	1.000	1.000	.999	.998
28	1.000	1.000	1.000	1.000	1.000	.999
29	1.000	1.000	1.000	1.000	1.000	1.000

الجدول (IV) : المساحة تحت المنحنى الطبيعي المعياري :

	Second decimal place in z									
z	0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01	0.00
−3.9										0.0000†
−3.8	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
−3.7	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
−3.6	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002	0.0002
−3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
−3.4	0.0002	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003
−3.3	0.0003	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0005	0.0005	0.0005
−3.2	0.0005	0.0005	0.0005	0.0006	0.0006	0.0006	0.0006	0.0006	0.0007	0.0007
−3.1	0.0007	0.0007	0.0008	0.0008	0.0008	0.0008	0.0009	0.0009	0.0009	0.0010
−3.0	0.0010	0.0010	0.0011	0.0011	0.0011	0.0012	0.0012	0.0013	0.0013	0.0013
−2.9	0.0014	0.0014	0.0015	0.0015	0.0016	0.0016	0.0017	0.0018	0.0018	0.0019
−2.8	0.0019	0.0020	0.0021	0.0021	0.0022	0.0023	0.0023	0.0024	0.0025	0.0026
−2.7	0.0026	0.0027	0.0028	0.0029	0.0030	0.0031	0.0032	0.0033	0.0034	0.0035
−2.6	0.0036	0.0037	0.0038	0.0039	0.0040	0.0041	0.0043	0.0044	0.0045	0.0047
−2.5	0.0048	0.0049	0.0051	0.0052	0.0054	0.0055	0.0057	0.0059	0.0060	0.0062
−2.4	0.0064	0.0066	0.0068	0.0069	0.0071	0.0073	0.0075	0.0078	0.0080	0.0082
−2.3	0.0084	0.0087	0.0089	0.0091	0.0094	0.0096	0.0099	0.0102	0.0104	0.0107
−2.2	0.0110	0.0113	0.0116	0.0119	0.0122	0.0125	0.0129	0.0132	0.0136	0.0139
−2.1	0.0143	0.0146	0.0150	0.0154	0.0158	0.0162	0.0166	0.0170	0.0174	0.0179
−2.0	0.0183	0.0188	0.0192	0.0197	0.0202	0.0207	0.0212	0.0217	0.0222	0.0228
−1.9	0.0233	0.0239	0.0244	0.0250	0.0256	0.0262	0.0268	0.0274	0.0281	0.0287
−1.8	0.0294	0.0301	0.0307	0.0314	0.0322	0.0329	0.0336	0.0344	0.0351	0.0359
−1.7	0.0367	0.0375	0.0384	0.0392	0.0401	0.0409	0.0418	0.0427	0.0436	0.0446
−1.6	0.0455	0.0465	0.0475	0.0485	0.0495	0.0505	0.0516	0.0526	0.0537	0.0548
−1.5	0.0559	0.0571	0.0582	0.0594	0.0606	0.0618	0.0630	0.0643	0.0655	0.0668
−1.4	0.0681	0.0694	0.0708	0.0721	0.0735	0.0749	0.0764	0.0778	0.0793	0.0808
−1.3	0.0823	0.0838	0.0853	0.0869	0.0885	0.0901	0.0918	0.0934	0.0951	0.0968
−1.2	0.0985	0.1003	0.1020	0.1038	0.1056	0.1075	0.1093	0.1112	0.1131	0.1151
−1.1	0.1170	0.1190	0.1210	0.1230	0.1251	0.1271	0.1292	0.1314	0.1335	0.1357
−1.0	0.1379	0.1401	0.1423	0.1446	0.1469	0.1492	0.1515	0.1539	0.1562	0.1587
−0.9	0.1611	0.1635	0.1660	0.1685	0.1711	0.1736	0.1762	0.1788	0.1814	0.1841
−0.8	0.1867	0.1894	0.1922	0.1949	0.1977	0.2005	0.2033	0.2061	0.2090	0.2119
−0.7	0.2148	0.2177	0.2206	0.2236	0.2266	0.2296	0.2327	0.2358	0.2389	0.2420
−0.6	0.2451	0.2483	0.2514	0.2546	0.2578	0.2611	0.2643	0.2676	0.2709	0.2743
−0.5	0.2776	0.2810	0.2843	0.2877	0.2912	0.2946	0.2981	0.3015	0.3050	0.3085
−0.4	0.3121	0.3156	0.3192	0.3228	0.3264	0.3300	0.3336	0.3372	0.3409	0.3446
−0.3	0.3483	0.3520	0.3557	0.3594	0.3632	0.3669	0.3707	0.3745	0.3783	0.3821
−0.2	0.3859	0.3897	0.3936	0.3974	0.4013	0.4052	0.4090	0.4129	0.4168	0.4207
−0.1	0.4247	0.4286	0.4325	0.4364	0.4404	0.4443	0.4483	0.4522	0.4562	0.4602
−0.0	0.4641	0.4681	0.4721	0.4761	0.4801	0.4840	0.4880	0.4920	0.4960	0.5000

† المساحة إلى يمين $z = 3.90$ هي 1 تقريباً.

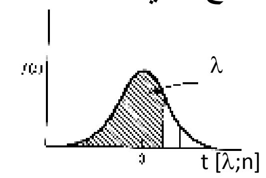
الجدول (IV) : المساحة تحت المنحنى الطبيعي المعياري.



	Second decimal place in z									
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000†									

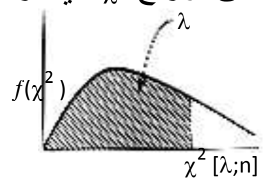
المساحة إلى يسار z - 3.90 هي تقريباً صفر.

الجدول (V) : يعطـي القيم علـى المحـور الأفقـي t وهـي : t [λ ; n] التـي يكـون إلـى يسارها λ من المساحة تحت توزيع t ذي درجات الحرية n.



			Probability (*p*)				
df	**.75**	**.90**	**.95**	**.975**	**.99**	**.995**	**.999**
1	1.000	3.078	6.314	12.706	31.821	63.657	318.31
2	.816	1.886	2.920	4.303	6.965	9.925	22.326
3	.765	1.638	2.353	3.182	4.541	5.841	10.213
4	.741	1.533	2.132	2.776	3.747	4.604	7.173
5	.727	1.476	2.015	2.571	3.365	4.032	5.893
6	.718	1.440	1.943	2.447	3.143	3.707	5.208
7	.711	1.415	1.895	2.365	2.998	3.499	4.785
8	.706	1.397	1.860	2.306	2.896	3.355	4.501
9	.703	1.383	1.833	2.262	2.821	3.250	4.297
10	.700	1.372	1.812	2.228	2.764	3.169	4.144
11	.697	1.363	1.796	2.201	2.718	3.106	4.025
12	.695	1.356	1.782	2.179	2.681	3.055	3.930
13	.694	1.350	1.771	2.160	2.650	3.012	3.852
14	.692	1.345	1.761	2.145	2.624	2.977	3.787
15	.691	1.341	1.753	2.131	2.602	2.947	3.733
16	.690	1.337	1.746	2.120	2.583	2.921	3.686
17	.689	1.333	1.740	2.110	2.567	2.898	3.646
18	.688	1.330	1.734	2.101	2.552	2.878	3.610
19	.688	1.328	1.729	2.093	2.539	2.861	3.579
20	.687	1.325	1.725	2.086	2.528	2.845	3.552
21	.686	1.323	1.721	2.080	2.518	2.831	3.527
22	.686	1.321	1.717	2.074	2.508	2.819	3.505
23	.685	1.319	1.714	2.069	2.500	2.807	3.485
24	.685	1.318	1.711	2.064	2.492	2.797	3.467
25	.684	1.316	1.708	2.060	2.485	2.787	3.450
26	.684	1.315	1.706	2.056	2.479	2.779	3.435
27	.684	1.314	1.703	2.052	2.473	2.771	3.421
28	.683	1.313	1.701	2.048	2.467	2.763	3.408
29	.683	1.311	1.699	2.045	2.462	2.756	3.396
30	.683	1.310	1.697	2.042	2.457	2.750	3.385
40	.681	1.303	1.684	2.021	2.423	2.704	3.307
60	.679	1.296	1.671	2.000	2.390	2.660	3.232
120	.677	1.289	1.658	1.980	2.358	2.617	3.160
∞	.674	1.282	1.645	1.960	2.326	2.576	3.090

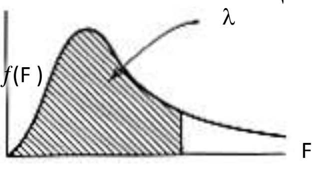
الجدول (VI) : يعطي القيم على المحور الأفقي χ^2 وهي $\chi^2 [\lambda ; n]$ التي يكون إلى يسارها λ من المساحة تحت منحنى توزيع χ^2 ذي درجات الحرية n.



Probability (p)

n	.01	.02	.05	.10	.20	.30	.50	.70	.80	.90	.95	.98	.99	.999
1	$.0^3157$	$.0^3628$	.00393	.0158	.0642	.148	.455	1.074	1.642	2.706	3.841	5.412	6.635	10.827
2	.0201	.0404	.103	.211	.446	.713	1.386	2.408	3.219	4.605	5.991	7.824	9.210	13.815
3	.115	.185	.352	.584	1.005	1.424	2.366	3.665	4.642	6.251	7.815	9.837	11.341	16.268
4	.297	.429	.711	1.064	1.649	2.195	3.357	4.878	5.989	7.779	9.488	11.668	13.277	18.465
5	.554	.752	1.145	1.610	2.343	3.000	4.351	6.064	7.289	9.236	11.070	13.388	15.086	20.517
6	.872	1.134	1.635	2.204	3.070	3.828	5.348	7.231	8.558	10.645	12.592	15.033	16.812	22.457
7	1.239	1.564	2.167	2.833	3.822	4.671	6.346	8.383	9.803	12.017	14.067	16.622	18.475	24.322
8	1.646	2.032	2.733	3.490	4.594	5.527	7.344	9.524	11.030	13.362	15.507	18.168	20.090	26.125
9	2.088	2.532	3.325	4.168	5.380	6.393	8.343	10.656	12.242	14.684	16.919	19.679	21.666	27.877
10	2.558	3.059	3.940	4.865	6.179	7.267	9.342	11.781	13.442	15.987	18.307	21.161	23.209	29.588
11	3.053	3.609	4.575	5.578	6.989	8.148	10.341	12.899	14.631	17.275	19.675	22.618	24.725	31.264
12	3.571	4.178	5.226	6.304	7.807	9.034	11.340	14.011	15.812	18.549	21.026	24.054	26.217	32.909
13	4.107	4.765	5.892	7.042	8.634	9.926	12.340	15.119	16.985	19.812	22.362	25.472	27.688	34.528
14	4.660	5.368	6.571	7.790	9.467	10.821	13.339	16.222	18.151	21.064	23.685	26.873	29.141	36.123
15	5.229	5.985	7.261	8.547	10.307	11.721	14.339	17.322	19.311	22.307	24.996	28.259	30.578	37.697
16	5.812	6.614	7.962	9.312	11.152	12.624	15.338	18.418	20.465	23.542	26.296	29.633	32.000	39.252
17	6.408	7.255	8.672	10.085	12.002	13.531	16.338	19.511	21.615	24.769	27.587	30.995	33.409	40.790
18	7.015	7.906	9.390	10.865	12.857	14.440	17.338	20.601	22.760	25.989	28.869	32.346	34.805	42.312
19	7.633	8.567	10.117	11.651	13.716	15.352	18.338	21.689	23.900	27.204	30.144	33.687	36.191	43.820
20	8.260	9.237	10.851	12.443	14.578	16.266	19.337	22.775	25.038	28.412	31.410	35.020	37.566	45.315
21	8.897	9.915	11.591	13.240	15.445	17.182	20.337	23.858	26.171	29.615	32.671	36.343	38.932	46.797
22	9.542	10.600	12.338	14.041	16.314	18.101	21.337	24.939	27.301	30.813	33.924	37.659	40.289	48.268
23	10.196	11.293	13.091	14.848	17.187	19.021	22.337	26.018	28.429	32.007	35.172	38.968	41.638	49.728
24	10.856	11.992	13.848	15.659	18.062	19.943	23.337	27.096	29.553	33.196	36.415	40.270	42.980	51.179
25	11.524	12.697	14.611	16.473	18.940	20.867	24.337	28.172	30.675	34.382	37.652	41.566	44.314	52.620
26	12.198	13.409	15.379	17.292	19.820	21.792	25.336	29.246	31.795	35.563	38.885	42.856	45.642	54.052
27	12.879	14.125	16.151	18.114	20.703	22.719	26.336	30.319	32.912	36.741	40.113	44.140	46.963	55.476
28	13.565	14.847	16.928	18.939	21.588	23.647	27.336	31.391	34.027	37.916	41.337	45.419	48.278	56.893
29	14.256	15.574	17.708	19.768	22.475	24.577	28.336	32.461	35.139	39.087	42.557	46.693	49.588	58.302
30	14.953	16.306	18.493	20.599	23.364	25.508	29.336	33.530	36.250	40.256	43.773	47.962	50.892	59.703

الجدول (VII) : يعطي القيم على المحور الأفقي F أي $F[\lambda; n_1, n_2]$ التي يكون إلى يسارها λ من المساحة تحت منحنى توزيع F ذي n_1 درجات حرية في البسط، n_2 درجات حرية في المقام.



$F[\lambda; n_1, n_2]$

1. PROBABILITY LEVEL $p = .95$

n_2 \ n_1	1	2	3	4	5	6	7	8	10	20	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	241.9	248.0	254.00
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.40	19.45	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.79	8.66	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	5.96	5.80	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.74	4.56	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.06	3.87	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.64	3.44	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.35	3.15	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.14	2.94	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	2.98	2.77	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.85	2.65	2.40
12	4.76	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.75	2.54	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.67	2.46	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.60	2.39	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.54	2.33	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.49	2.28	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.45	2.23	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.41	2.19	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.64	2.48	2.38	2.16	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.35	2.12	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.32	2.10	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.30	2.07	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.27	2.05	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.25	2.03	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.24	2.01	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.22	1.99	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.20	1.97	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.19	1.96	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.18	1.94	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.16	1.93	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.08	1.84	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	1.99	1.75	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.91	1.66	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.83	1.57	1.00

تكملة جدول (VII)

2. PROBABILITY LEVEL $\lambda = 0.975$

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	20	∞
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.6	993.1	1018
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.45	39.50
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.17	13.90
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.56	8.26
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.33	6.02
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.17	4.85
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.47	4.14
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.00	3.67
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.67	3.33
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.42	3.08
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.23	2.88
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.07	2.72
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	2.95	2.60
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	2.84	2.49
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.76	2.40
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.68	2.32
17	6.04	4.62	4.01	3.66	3.44	3.29	3.16	3.06	2.98	2.92	2.62	2.25
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.56	2.19
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.51	2.13
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.46	2.09
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.42	2.04
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.39	2.00
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.36	1.97
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.33	1.94
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.30	1.91
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	2.59	2.28	1.88
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63	2.57	2.25	1.85
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55	2.23	1.83
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59	2.53	2.21	1.81
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.20	1.79
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.07	1.64
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	1.94	1.48
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	1.82	1.31
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.06	1.71	1.00

تكملة جدول (VII)

2. PROBABILITY LEVEL $\lambda = 0.99$

n_2 \ n_1	1	2	3	4	5	6	7	8	10	20	∞
1	4052	4999.5	5403	5625	5764	5859	5928	5981	6056	6209	6366
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.40	99.45	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.23	26.69	26.12
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.55	14.02	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.05	9.55	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.87	7.40	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.62	6.16	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.81	5.36	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.26	4.81	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.85	4.41	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.54	4.10	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.30	3.86	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.10	3.66	3.16
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	3.94	3.51	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.80	3.37	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.69	3.26	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.59	3.16	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.51	3.08	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.43	3.00	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.37	2.94	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.31	2.88	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.26	2.83	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.21	2.78	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.17	2.74	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.13	2.70	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.09	2.66	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.06	2.63	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.03	2.60	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.00	2.57	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	2.98	2.55	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.80	2.37	1.81
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.63	2.20	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.47	2.03	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.32	1.88	1.00

الجدول (VIII) : القيم الطبيعية Normal scores.

Normal Scores

الترتيب Ordered position	n								
	5	6	7	8	9	10	11	12	13
1	−1.18	−1.28	−1.36	−1.43	−1.50	−1.55	−1.59	−1.64	−1.68
2	−0.50	−0.64	−0.76	−0.85	−0.93	−1.00	−1.06	−1.11	−1.16
3	0.00	−0.20	−0.35	−0.47	−0.57	−0.65	−0.73	−0.79	−0.85
4	0.50	0.20	0.00	−0.15	−0.27	−0.37	−0.46	−0.53	−0.60
5	1.18	0.64	0.35	0.15	0.00	−0.12	−0.22	−0.31	−0.39
6		1.28	0.76	0.47	0.27	0.12	0.00	−0.10	−0.19
7			1.36	0.85	0.57	0.37	0.22	0.10	0.00
8				1.43	0.93	0.65	0.46	0.31	0.19
9					1.50	1.00	0.73	0.53	0.39
10						1.55	1.06	0.79	0.60
11							1.59	1.11	0.85
12								1.64	1.16
13									1.68

الترتيب Ordered position	n								
	14	15	16	17	18	19	20	21	22
1	−1.71	−1.74	−1.77	−1.80	−1.82	−1.85	−1.87	−1.89	−1.91
2	−1.20	−1.24	−1.28	−1.32	−1.35	−1.38	−1.40	−1.43	−1.45
3	−0.90	−0.94	−0.99	−1.03	−1.06	−1.10	−1.13	−1.16	−1.18
4	−0.66	−0.71	−0.76	−0.80	−0.84	−0.88	−0.92	−0.95	−0.98
5	−0.45	−0.51	−0.57	−0.62	−0.66	−0.70	−0.74	−0.78	−0.81
6	−0.27	−0.33	−0.39	−0.45	−0.50	−0.54	−0.59	−0.63	−0.66
7	−0.09	−0.16	−0.23	−0.29	−0.35	−0.40	−0.45	−0.49	−0.53
8	0.09	0.00	−0.08	−0.15	−0.21	−0.26	−0.31	−0.36	−0.40
9	0.27	0.16	0.08	0.00	−0.07	−0.13	−0.19	−0.24	−0.28
10	0.45	0.33	0.23	0.15	0.07	0.00	−0.06	−0.12	−0.17
11	0.66	0.51	0.39	0.29	0.21	0.13	0.06	0.00	−0.06
12	0.90	0.71	0.57	0.45	0.35	0.26	0.19	0.12	0.06
13	1.20	0.94	0.76	0.62	0.50	0.40	0.31	0.24	0.17
14	1.71	1.24	0.99	0.80	0.66	0.54	0.45	0.36	0.28
15		1.74	1.28	1.03	0.84	0.70	0.59	0.49	0.40
16			1.77	1.32	1.06	0.88	0.74	0.63	0.53
17				1.80	1.35	1.10	0.92	0.78	0.66
18					1.82	1.38	1.13	0.95	0.81
19						1.85	1.40	1.16	0.98
20							1.87	1.43	1.18
21								1.89	1.45
22									1.91

تكملة الجدول (VIII).

الترتيب Ordered position	n							
	23	24	25	26	27	28	29	30
1	−1.93	−1.95	−1.97	−1.98	−2.00	−2.01	−2.03	−2.04
2	−1.48	−1.50	−1.52	−1.54	−1.56	−1.58	−1.59	−1.61
3	−1.21	−1.24	−1.26	−1.28	−1.30	−1.32	−1.34	−1.36
4	−1.01	−1.04	−1.06	−1.09	−1.11	−1.13	−1.15	−1.17
5	−0.84	−0.87	−0.90	−0.93	−0.95	−0.98	−1.00	−1.02
6	−0.70	−0.73	−0.76	−0.79	−0.82	−0.84	−0.87	−0.89
7	−0.57	−0.60	−0.63	−0.66	−0.69	−0.72	−0.75	−0.77
8	−0.44	−0.48	−0.52	−0.55	−0.58	−0.61	−0.64	−0.67
9	−0.33	−0.37	−0.41	−0.44	−0.48	−0.51	−0.54	−0.57
10	−0.22	−0.26	−0.30	−0.34	−0.38	−0.41	−0.44	−0.47
11	−0.11	−0.15	−0.20	−0.24	−0.28	−0.31	−0.35	−0.38
12	0.00	−0.05	−0.10	−0.14	−0.18	−0.22	−0.26	−0.29
13	0.11	0.05	0.00	−0.05	−0.09	−0.13	−0.17	−0.21
14	0.22	0.15	0.10	0.05	0.00	−0.04	−0.09	−0.12
15	0.33	0.26	0.20	0.14	0.09	0.04	0.00	−0.04
16	0.44	0.37	0.30	0.24	0.18	0.13	0.09	0.04
17	0.57	0.48	0.41	0.34	0.28	0.22	0.17	0.12
18	0.70	0.60	0.52	0.44	0.38	0.31	0.26	0.21
19	0.84	0.73	0.63	0.55	0.48	0.41	0.35	0.29
20	1.01	0.87	0.76	0.66	0.58	0.51	0.44	0.38
21	1.21	1.04	0.90	0.79	0.69	0.61	0.54	0.47
22	1.48	1.24	1.06	0.93	0.82	0.72	0.64	0.57
23	1.93	1.50	1.26	1.09	0.95	0.84	0.75	0.67
24		1.95	1.52	1.28	1.11	0.98	0.87	0.77
25			1.97	1.54	1.30	1.13	1.00	0.89
26				1.98	1.56	1.32	1.15	1.02
27					2.00	1.58	1.34	1.17
28						2.01	1.59	1.36
29							2.03	1.61
30								2.04

الجدول (IX) : تحويل r إلى فيشر z .

Transformation of r to Z $\left(\textbf{i.e.,}\ \ Z = \frac{1}{2}\ln\frac{1+r}{1-r} \right)$

r	Z	r	Z	r	Z
.00	.000				
.01	.010	.36	.377	.71	.887
.02	.020	.37	.388	.72	.908
.03	.030	.38	.400	.73	.929
.04	.040	.39	.412	.74	.950
.05	.050	.40	.424	.75	.973
.06	.060	.41	.436	.76	.996
.07	.070	.42	.448	.77	1.020
.08	.080	.43	.460	.78	1.045
.09	.090	.44	.472	.79	1.071
.10	.100	.45	.485	.80	1.099
.11	.110	.46	.497	.81	1.127
.12	.121	.47	.510	.82	1.157
.13	.131	.48	.523	.83	1.188
.14	.141	.49	.536	.84	1.221
.15	.151	.50	.549	.85	1.256
.16	.161	.51	.563	.86	1.293
.17	.172	.52	.576	.87	1.333
.18	.182	.53	.590	.88	1.376
.19	.192	.54	.604	.89	1.422
.20	.203	.55	.618	.90	1.472
.21	.213	.56	.633	.91	1.528
.22	.224	.57	.648	.92	1.589
.23	.234	.58	.662	.93	1.658
.24	.245	.59	.678	.94	1.738
.25	.255	.60	.693	.95	1.832
.26	.266	.61	.709	.96	1.946
.27	.277	.62	.725	.97	2.092
.28	.288	.63	.741	.98	2.298
.29	.299	.64	.758	.99	2.647
.30	.310	.65	.775		
.31	.321	.66	.793		
.32	.332	.67	.811		
.33	.343	.68	.829		
.34	.354	.69	.848		
.35	.365	.70	.867		